STRUCTURE AND LIGAND-BASED APPLICATIONS OF MOLECULAR MODELING TO GAIN INSIGHTS INTO THE STRUCTURAL FEATURES OF PROTEINS AND SMALL MOLECULES OF BIOLOGICAL INTEREST

by

VASUDEVAN CHANDRASEKARAN
(Under the Direction of Ethan Will Taylor)

ABSTRACT

Computational and experimental methodologies are often seen as competing resources. This is unfortunate because in the evaluation of complex biological phenomena often both methodologies are required to develop a complete understanding of the underlying science as well as an understanding of the system level behavior. Computational modeling approaches have become important in small molecule as well as macromolecular studies, especially when the problems are not readily accessible by experimental methods or when the computational methods can supplement the results obtained from experimental techniques and provide a basis for interpretation of the data.

This dissertation reports computational studies of choline acetyltransferase inhibitors, the oxidized form of nuclear factor-kappa B homodimer and the complex structure of vascular endothelial growth factor (VEGF) and SPARC. A variety of ligand and structure-based molecular modeling techniques have been used throughout the dissertation research. Work investigating choline acetyltransferase inhibitors employed ligand-based methods like 3D-QSAR to explore trans-1-methyl-4-(1-naphthylvinyl) pyridinium (MNVP⁺) analogs and predict structural features that are essential for selective and potent inhibition. The information on the molecules sharing a given activity and their most similar analogs that do not exhibit this activity was used to develop models that can be used to discriminate between active and inactive compounds prior to their synthesis.

The question of structure, function and nuclear entry of the oxidized form of NF-κB homodimer was explored through protein modeling and molecular dynamics methods. It is our hypothesis that after I-κB degradation the two subunits of the NF-κB homodimer adopt a closed conformation through a hinge movement involving the flexible linker region, making the dimer
structure compact to facilitate its nuclear translocation. The results from molecular modeling studies help to determine how the disulfide bridge formation between the two subunits of the dimer facilitates nuclear entry, through comparison with the experimentally known structure of the reduced form.

The structure of the complex between vascular endothelial growth factor (VEGF) with anti-angiogenic molecule, SPARC (secreted protein acidic and rich in cysteine) is predicted through protein-protein docking simulation used in conjunction with biological data. Docking method was used to model the complex structure, in an effort to identify the specific interactions between the two molecules, which can be used as a basis for guiding future studies aimed at identifying inhibitors for VEGF-induced angiogenesis.

INDEX WORDS: Molecular Modeling, Computer-aided drug discovery, 3D-QSAR, Choline acetyltransferase inhibitors, Nuclear Factor-κB, Thioredoxin, Protein-Protein docking, Angiogenesis, Vascular Endothelial Growth Factor, SPARC
STRUCTURE AND LIGAND-BASED APPLICATIONS OF MOLECULAR MODELING TO
GAIN INSIGHTS INTO THE STRUCTURAL FEATURES OF PROTEINS AND SMALL
MOLECULES OF BIOLOGICAL INTEREST

by

VASUDEVAN CHANDRASEKARAN
B.Pharm., Birla Institute of Technology and Science, Pilani, India, 2002

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2006
STRUCTURE AND LIGAND-BASED APPLICATIONS OF MOLECULAR MODELING TO GAIN INSIGHTS INTO THE STRUCTURAL FEATURES OF PROTEINS AND SMALL MOLECULES OF BIOLOGICAL INTEREST

by

VASUDEVAN CHANDRASEKARAN

Major Professor: Ethan Will Taylor
Committee: Phillip Bowen, Diane Hartle, Warren Beach, Anthony Capomacchia

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2006
DEDICATION

To my parents,

Who encouraged my academic pursuits and supported throughout the past years.
ACKNOWLEDGEMENTS

This has by no means been an individual effort and there are many people to thank. First of all, working for Dr. Taylor has been fantastic. He has been a constant source of encouragement and support throughout the whole experience. He gave me every freedom to work independently on problems and pursue research ideas. Not only was he there to guide me in my research, but he was always there to help me in other situations in the last four years of graduate school. He has been a perfect advisor and I thank him for his patient help, insights and interesting discussions about various other things.

I would also like to thank Dr. Bowen for all the help, support and guidance he has given over the years. I am grateful for the opportunities he provided for me to gain research skills, his constant encouragement and his belief in my work. I am deeply indebted to both of them, and without them this would not have been possible.

My gratitude is also extended to the members of my advisory committee, Drs. Diane Hartle, Warren Beach and Anthony Capomacchia for their time, expert guidance and critiques which have been an integral part of my education. I would especially like to thank Dr. Hartle for her time and effort in helping me on several occasions and providing valuable suggestions.

The graduate program at UGA was made memorable by my friends. Many thanks to all of you being there and brightening up my days. The experience would not have been the same without all of you. Finally, I would like to acknowledge everyone who has contributed directly or indirectly to the completion of this research.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ............................................................................................................. v

LIST OF TABLES ....................................................................................................................... viii

LIST OF FIGURES ..................................................................................................................... ix

CHAPTER

1 INTRODUCTION AND LITERATURE REVIEW ................................................................. 1

1.1 CHALLENGE OF DRUG DISCOVERY .......................................................................... 1

1.2 COMPUTER-AIDED DRUG DISCOVERY .................................................................... 2

1.3 LIGAND-BASED APPROACH ..................................................................................... 5

1.4 STRUCTURE-BASED APPROACH ............................................................................... 10

1.5 OVERVIEW OF CHAPTERS ....................................................................................... 16

1.6 REFERENCES .............................................................................................................. 21

2 THREE DIMENSIONAL QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (3D-QSAR) ANALYSES OF CHOLINE ACETYLTRANSFERASE INHIBITORS .............................................................. 30

2.1 ABSTRACT ..................................................................................................................... 31

2.2 INTRODUCTION .......................................................................................................... 32

2.3 MATERIALS AND METHODS .................................................................................... 35

2.4 RESULTS AND DISCUSSION ..................................................................................... 40

2.5 CONCLUSION .............................................................................................................. 49
LIST OF TABLES

Table 2.1: Chat inhibitory activity and Moriguchi LogP values used in the CoMFA and CoMSIA study .........................................................................................................................38
Table 2.2: PLS statistical data obtained from the CoMFA and CoMSIA models ..................42
Table 2.3: Results of the predicted activities and residuals of the compounds by the CoMFA and CoMSIA models ........................................................................................................43
Table 2.4: Chat inhibitory activity and Moriguchi LogP values of the test set molecules ..........48
Table 2.5: Predicted activities and residuals of the test set by the CoMSIA model ..............49
Table 4.1: Residues at the interface responsible for hydrogen-bonding interactions ...........91
Table 4.2: Residues at the interface responsible for hydrophobic interactions ....................91
Table 4.3: Results of the predicted contributions of residues through virtual alanine scanning ...94
Table 4.4: Results of the predicted contributions of residues through virtual alanine scanning ...94
LIST OF FIGURES

Figure 1.1: Sample 3D Pharmacophore ................................................................. 6
Figure 1.2: QSAR Methodology .............................................................................. 8
Figure 1.3: Graphical Display of CoMFA Electrostatic Field ................................. 10
Figure 1.4: Structural Prediction by Homology Modeling ........................................ 12
Figure 1.5: Interaction Terms in Potential Energy Function ............................... 15
Figure 2.1: trans-1-methyl-4-(1-naphthylvinyl)pyridinium ................................. 34
Figure 2.2: Alignment of ChAT inhibitors using common substructure method ...... 37
Figure 2.3: CoMFA steric contour plot ................................................................. 43
Figure 2.4: CoMFA electrostatic contour plot ......................................................... 44
Figure 2.5: CoMSIA hydrophobic contour plot ...................................................... 45
Figure 2.6: CoMSIA H-bond donor contour plot .................................................... 46
Figure 2.7: CoMSIA H-bond acceptor contour plot ............................................ 46
Figure 2.8: Comparison of actual vs. predicted –log IC$_{50}$ of the molecules .......... 47
Figure 3.1: Three-dimensional structure of NF-κB p50 homodimer with the DNA removed .......... 60
Figure 3.2: Superposition of the NF-κB dimers before and after hinge movement .......... 61
Figure 3.3: Close-up view of the hinge region residues before and after torsional modification .... 65
Figure 3.4: Ramachandran plot for the oxidized form of NF-κB ............................... 67
Figure 3.5: Space-filling model of the oxidized and reduced forms of NF-κB ............. 69
Figure 3.6: Comparison of the DNA-binding cleft of the reduced and oxidized forms ........ 70
Figure 4.1: Protein-Protein Docking Procedure ...................................................... 83
Figure 4.2: Complex structure of VEGF-SPARC obtained through docking simulation ..........88

Figure 4.3: VEGF receptor binding interactions.................................................................90

Figure 4.4: Amino acid residues involved in side chain interactions at the VEGF-SPARC binding interface ..................................................................................................................92
CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

1.1 Challenge of Drug Discovery

The drug discovery process is a long, painstaking and an expensive endeavor. Despite advances in technology and understanding of biological systems, drug discovery is still a long process with low rate of new therapeutic discovery. Only one compound of 10,000 makes it to the market and commercialization of a new drug today requires an average investment of $880 million and 15 years of development, including the cost and time to discover potential biological targets. Out of the ~15 years in development time of a successful compound, about 6 years are devoted to the drug discovery and the preclinical phase, 6.7 years to clinical trials and 2.2 years to the approval phase.

All drugs that are presently on the market are estimated to target less than 500 biomolecules, ranging from nucleic acids to enzymes, G-protein-coupled receptors (GPCRs) and ion channels. The choice of targets and approach used to discover hits for these targets depends on the chemical tractability of targets for small molecule intervention. It was thought that the availability of information on human genome, its sequence and what it encodes, would overcome the bottleneck of identifying new therapeutic targets which has been one limiting factor on the rate of new drug discovery. However, data indicates that new targets as opposed to established targets are more prone to drug discovery project failure in general. Irrespective of the target selected, the outcome of the hit identification process depends largely on the scientific
understanding of the target and the approach and technology used. The amount of background information available on the target, in particular its functional information, of how the target works both in normal physiology and how it is involved in human pathology is very critical in drug discovery. With increasing computer power becoming available at lower and lower costs, there has been a large focus towards developing computational methods to aid in the ever complex and multi-step discovery process.

1.2 Computer-Aided Drug Discovery

1.2.1 Need for Computational Approaches

Why do we need a computational approach to drug discovery….why cannot we use traditional approaches or newer methods like combinatorial library design, high-throughput screening? A retrospective analysis of the past discovery programs show that technologies like combinatorial library design and high-throughput screening (HTS) alone have demonstrated fewer success rates than anticipated. Although there has been a lot of advances and investments made in these technologies, recent reports have shown that they have not produced significant number of hits.\(^6\) Hundreds of thousands to millions of compounds have to be evaluated for a typical high-throughput search, without any knowledge of an active lead, and still there is no guarantee of success.\(^7,8\) Not all targets are suitable for HTS screening and in order to adapt the assays for high-throughput systems, technical and biological compromises have to be made which leads to high error rates. Even when compounds are identified by HTS they are not always suitable for initiation of further medicinal chemistry exploration towards optimizing and validating their drug-like properties.\(^9\) A large number of compounds synthesized did not result in the increase in drug candidates that was originally expected, so the focus of drug discovery requires a more targeted approach rather than just massive random screening and brute force approaches.
Efforts towards exploiting the structural information of the targets as well as the ligands are important for a successful drug research program. The ability to access information contained in biological macromolecules has brought about a better understanding of the protein structure and its active sites and the way molecules would interact with these target structures - thereby greatly enhancing the ability to design more efficient drugs using this information. This rational approach to drug design is used in the biopharmaceutical industry to discover and develop new drug compounds.

Drug discovery has often been compared to a search for a needle in a haystack. Knowledge of the detailed atomic three-dimensional structure of the protein and its ligand complexes should facilitate the design of novel, high affinity molecules that interact with that protein. This approach will help in reducing the size of the haystack. Computational modeling plays an important role in the process of elucidating the atomic structure of proteins and their complexes, and the design of novel, therapeutically relevant ligands based on these structural data. The computational chemist in a drug discovery setting plays an interactive role, acquiring and providing information among many scientists in various fields.

1.2.2 Experimental and Computational Methods

Why can we not simply use experimental 3D structures as they become available and why is molecular modeling necessary? Experimental approaches like X-ray crystallography, NMR and Cryo-EM provide valuable information about the 3D structure of macromolecules but there are certain limitations associated with each of these experimental approaches. Structural experimental data on molecules is rarely complete, for example, some critical side chains are poorly resolved, hypervariable loops are not resolved, and sometimes cleaved out to provide a more soluble protein for crystallization. Hydrogen atoms which are central to much enzymatic
activity are rarely resolved, metal ions are sometimes missing and active sites and/or exosites are blocked with inhibitors so that crystallization of enzymatically active forms is possible. Sometimes certain domains are missing, and most importantly, a crystal structure represents only a static average snapshot in a pseudo-solid environment with little clue of the dynamic character of the protein in a solvated, ionic medium.

It is technically very challenging to make crystals of proteins. Structures of protein-protein complexes, heterogeneous samples and membrane proteins are especially difficult to determine (cryo-EM methods are promising). Therefore computational modeling approaches have become especially important and are complementary to the experimental structure determination methods. Structure-based modeling, including solvent-inclusive dynamics and docking methods, offer an accurate approach of projecting-out the limitations of the structural data. So the answer to the posed question is that one does wish to employ new structural data as it becomes available, but in concert with complementary modeling/dynamics for a detailed understanding of the 3D structure of biomolecules. Molecular modeling and 3D structure-based design support the target selection and rational design of high-affinity protein ligands when integrated with experimental techniques like protein crystallography, NMR and high-throughput screening methods.

Rational drug design uses a variety of computational methods to design and identify novel compounds, solve biomolecular structures and understand molecular mechanisms. These methods fall into two major categories – ligand-based and structure-based drug design – depending on how much information is available about the potential drug compounds and the drug targets. Computational techniques can be potentially useful in elucidating the structure of biomolecules or in describing the interaction of protein-protein and protein-ligand complexes or
in predicting high-affinity ligands for a target. These different computational methods fall into two broad categories - ligand-based and structure-based approaches.

1.3 Ligand-based Approach

There are instances in drug design research when the only available information is the molecules sharing a given activity and their most similar analogues that do not exhibit this activity, which can be provided by biological testing or high-throughput screening. Also, many receptors are not readily amenable to receptor-based drug design. For example, many important receptors are membrane-bound proteins, which are notoriously difficult to crystallize. In such cases, a lead compound or active ligand must be found, and then the structure of the ligand guides the drug design process.

For a chemical series belonging to a common scaffold or related scaffolds, when a certain number active and inactive ligands for a target are already known, 2D or 3D similarity methods, QSAR or 3D-QSAR approaches and pharmacophore approaches can be used to derive structure-activity hypotheses.$^8$

1.3.1 Pharmacophore Model

For many proteins in the human body there is still no three-dimensional structure available. The ligand-based pharmacophore approach allows the activity prediction of molecules by means of common chemical features even if the bioactive binding conformation is not known. A pharmacophore is a specific, three dimensional map of biological properties common to all active conformations of a set of ligands which exhibit a particular activity. It is a conceptual template for drug design and once it is extracted from a set of ligands, a pharmacophore can be used as a model for the design of other molecules that can accomplish the same activity.
Modeling programs can be applied for automated ligand-based pharmacophore generation. The algorithms generate 3D pharmacophores based on alignment of common features present in highly potent compounds, by comparing a set of conformational models and a number of three-dimensional configurations of chemical features shared among the training set molecules. Pharmacophore features may be necessary for the activity, however, they are not sufficient and screening is needed to discriminate between active and inactive compounds. Based on the derived pharmacophore model, databases of real and virtual compounds can be searched to select screening compounds. Pharmacophore modeling methods can best be used when the structure of the macromolecular receptor is not available. However, pharmacophore modeling can be used as a complementary tool to standard techniques, such as docking, even when the three dimensional structure of the receptor is known.

\[\text{Aromatic Ring} \]
\[4.5-7.5 \text{ Angstroms} \]
\[7.6-8.5 \text{ Angstroms} \]
\[2-4 \text{ Angstroms} \]
\[\text{Hydrogen bond acceptor} \]
\[\text{Hydrogen Bond Donor} \]

**Figure 1.1: Sample 3D Pharmacophore**
1.3.2 Quantitative Structure-Activity Relationships

Quantitative structure activity relationships (QSARs or more general SARs) are based on a fundamental principle in chemistry, that chemical properties and the biological behavior elicited are a direct result of chemical structure. It is quantitative in that it employs a set of methods that tries to find a mathematical relationship between a set of descriptors (geometric, chemical etc.) of molecules and the observed activity.

Classical QSAR correlates biological activities of drugs with physicochemical properties or indicator variables that encode certain structural features.\textsuperscript{11-14} It attempts to find consistent relationship between biological activity and molecular properties, so that these rules can be used to evaluate the activity of new compounds. The first application of QSAR is attributed to Hansch (1969), who developed an equation that related biological activity to certain electronic characteristics and the hydrophobicity of a set of structures.\textsuperscript{15}

$$\log \left( \frac{1}{C} \right) = k_1 \log P - k_2 (\log P)^2 + k_3 \sigma + k_4$$

where;

- $C$ = minimum effective dose
- $P$ = octanol-water partition coefficient
- $\sigma$ = Hammett substituent constant
- $k_x$ = constants derived from regression analysis

Log P is a measure of the drug’s hydrophobicity, which was selected as a measure of its ability to pass through cell membranes. The log P (or log $P_{O/W}$) value reflects the relative solubility of the drug in octanol representing the lipid bi-layer of a cell membrane and water, the fluid within the cell and in blood. Tables of $\sigma$ values for numerous substituents have been
Statistical analysis is used to determine which of the variables best describe or correlate with the observed biological activity, and which ones are cross-correlated. The final QSAR involves only the most important 3 to 5 descriptors, eliminating those with high cross-correlation.

Figure 1.2: QSAR Methodology

Early QSAR analyses however did not consider the 3D descriptors or chirality of the molecules and were therefore a poor approximation of a molecule’s biological behavior. Three-dimensional QSAR methods were later developed that took into account the conformations and molecular fields of molecules. The important aims of these models are to correlate the three-dimensional structure of drug molecules with their biological activities and to be able to predict the activity of new molecules prior to synthesis. The rationale is that once a valid QSAR has been determined, it should be possible to predict the biological activity of related drug candidates before they are put through expensive and time-consuming biological testing.
3D-QSAR maps important regions of bioactive molecules in 3D space, such that regions of hydrophobicity, hydrophilicity, H-bond acceptor, H-bond donor etc. are rendered so that they overlap, and a general 3D pattern of the functionally significant parts of an active drug are determined. One of the most widely used tools for QSAR and drug design is Comparative Molecular Field Analysis (CoMFA).\textsuperscript{20,21} The idea underlying a CoMFA is that differences in a target property are often related to differences in the shapes of the non-covalent fields surrounding the tested molecules.

Ligand-based three-dimensional quantitative structure-activity relationship (3D-QSAR) methods, like the Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA),\textsuperscript{22,23} have supported the chemical optimization of numerous lead compounds.\textsuperscript{24-26} CoMSIA considers both the steric and electrostatic features, in addition to hydrogen bond donor, hydrogen bond acceptor and hydrophobic fields. These fields are sampled at the intersections of one or more lattices or grids spanning a 3D region around active molecules. A requirement for these methods is that all molecules have to interact with the same kind of receptor with identical binding sites in the same relative geometry. The QSAR equation is generated by a special Partial Least Square (PLS) analysis and evaluated by its cross-validated $r^2$ ($q^2$) value.

These QSAR methods capture the three-dimensional information concerning the ligands and their specific conformations, and are therefore more advanced techniques that attempt to model the receptor environment from the perspective of the ligands. Thus, CoMFA and CoMSIA studies allow chemical modifications that are beneficial or detrimental for the biological activity to be recognized. These models can also be used to recognize molecular features that are responsible for selectivity of the ligands.
The advantage of ligand-based methods is that the only information required is the molecules sharing a given activity and their most similar analogues that do not exhibit this activity. The hypothesis is that since all the information resides in these molecules, there is no need to know the receptor or the mechanism of action. However, this approach does not work well in cases where a detailed knowledge of the receptor structure is required, as computational processes sometimes cannot discriminate between active and inactive molecules, as well as the receptors.

1.4 Structure-based Approach

Structural data on proteins are important for gaining molecular understanding of the structures, their active sites and binding specificities, and they provide a natural starting point for designing therapeutic leads, suggesting mutation experiments and protein engineering. Massive amounts of protein sequence data are available from modern large-scale DNA sequencing efforts such as the Human Genome Project. Despite community-wide efforts in structural genomics, the output of
experimentally determined protein structures through X-ray crystallography or NMR spectroscopy is lagging far behind the output of protein sequences, as these experimental methods are time consuming in addition to being relatively expensive. With the advances made in computational methods, protein modeling is routinely used to overcome this problem by simulating protein folding and predicting the tertiary structure of proteins and their complexes.

1.4.1 Comparative Protein Modeling

Comparative or homology modeling involves predicting the three-dimensional structure of a given protein sequence (target) based on an alignment to one or more known protein structures (templates).

In general, structure is more conserved in nature than sequence and though the number of actual proteins is vast, there is a limited set of tertiary structural motifs to which most proteins belong. It is therefore expected that two proteins of similar origin and function would have reasonable structural similarity and it is possible to use the known structure as a template for modeling the structure of the unknown. Sequence identity of over 60% between the target and template leads to a good structure prediction and a sequence identity of at least 30% is required for generating reliable models from homology modeling. Homology modeling, in general, involves four steps: fold assignment, sequence alignment, model building and model refinement. The first step is to search a sequence database of proteins with known structures using sequence similarity search algorithms like BLAST and FASTA to identify known three-dimensional structures related to the target sequence. Once a significant similarity is found, the sequences are then aligned using single or multiple sequence alignments. Multiple sequence alignment usually improves the quality of the final structure prediction by considering further homologues.
After the known structures are aligned, they are examined to identify the structurally conserved regions (SCRs) from which an average structure or framework, can be constructed for these regions of the proteins. A model of the target protein is built by introducing insertions and deletions according to the sequence alignment.

Variable regions like loops are less conserved and can adopt different conformations in even closely different homologues. When generating coordinates for the unknown structure, the main chain atoms are modeled based on the known structure but the side chain conformations are either predicted from the known structure if the target sequence is identical or modeled based on a systematic exploration of the possible rotamers.

1.4.2. Model Refinement

Once the variable regions and side chains are modeled, the homology model is evaluated based on structural analysis of phi-psi angles, peptide bond planarity, bond lengths, bond angles, hydrogen-bond geometry, and side-chain conformations. For large molecular systems like
proteins which comprise thousands of atoms, modeling techniques which describe the potential energy surface in terms of quantum mechanics (QM) are too demanding in terms of computational time to be of much use. Molecular mechanics (MM) approaches are widely applied for such systems for molecular structure refinement, molecular dynamics simulations and docking simulations. MM techniques use classical mechanics to analyze the structure and dynamics of molecular systems and treats atoms as classical particles. Molecules are simplified as mechanical assemblies of elements such as balls, sticks and flexible joints and MM force fields apply terms to describe van der Waals, electrostatic and hydrogen-bonding interactions between the atoms. The potential function calculates the systems potential energy in a given conformation as the sum of the individual energy terms. The exact functional form of the potential function depends on the particular simulation program being used. This simplification allows this method to be applied to larger molecular systems that cannot be studied by ab initio methods.

The structure can be optimized by energy minimization using force field methods to remove bad steric clashes and improve the energetic and other criteria, by reducing the gradients and the energy of the system. Molecular dynamic (MD) simulations using explicit solvent can also be used to refine the model, as MD simulations attempt de novo prediction of relevant parameters from the system’s starting structure and are capable of reproducing experimentally accessible properties. Molecular dynamics is a computer simulation technique where the time evolution of a set of interacting atoms is followed by integrating their equations of motion. While an energy minimization calculation will find a local minimum in the potential energy surface, molecular dynamics calculations can cover a far broader sample of conformations. By giving each particle a velocity, molecular dynamics imparts kinetic energy to the system. This energy
can be sufficient to enable the system to progress over barriers in the potential surface which could not be crossed in a gradient minimization procedure. MD method accounts for both the intrinsic flexibility and explicit solvation of the system giving a dynamic, rather than static, picture of the biomolecular system. Typical MD simulations can be performed on systems containing tens of thousands to even millions of atoms, and for simulation times ranging from a few picoseconds to hundreds of nanoseconds. While these numbers are certainly impressive, there are definitely instances where time and/or size limitations of the system become important.

The accuracy of a protein simulation using MD techniques largely depends on the force field applied, in terms of how the force field treats different interactions like long range non-bonded interactions etc. and whether the most appropriate force field is selected for the system under investigation. Majority of the protein simulations are performed using all-atom force field model, including AMBER\textsuperscript{37} and CHARMM\textsuperscript{38} forcefields. Forcefields enable the potential energy of a molecular system to be calculated rapidly and fairly accurately, and the potential energy function is the core of any force field that describes the relationship of the structure to the energy of the system.\textsuperscript{39} A typical potential energy function used in biomolecular forcefields to calculate the forces on particles is shown in equations 1.1 and 1.2.

MD calculations evaluate the forces acting on each particle and use these to determine the accelerations of these particles. Since in MD simulations the molecular geometry is allowed to evolve naturally using the inter-atomic forces between the various constituents of the system, simulating the dynamics of proteins and understanding the molecular motions gives a more accurate representation of their structures and help determine their chemical and biological properties.
Intramolecular (internal, bonded terms)

\[ \sum_{bonds} K_b (b - b_o)^2 + \sum_{angles} K_\theta (\theta - \theta_o)^2 + \sum_{torsions} K_\phi (1 + \cos(n\phi - \delta)) \]

\[ + \sum_{impropers} K_\phi (\varphi - \varphi_o)^2 + \sum_{Urey-Bradley} K_{UB} (r_{1,3} - r_{1,3,o})^2 \]  
\[ (1.1) \]

Intermolecular (external, nonbonded terms)

\[ \sum_{nonbonded} \frac{q_i q_j}{4 \pi \varepsilon_0} + \frac{\varepsilon_{ij}}{4 \pi \varepsilon_0} \left[ \left( \frac{R_{min,ij}}{r_{ij}} \right)^6 - \left( \frac{R_{min,ij}}{r_{ij}} \right)^{12} \right] \] 
\[ (1.2) \]

\[ V_{bond} = K_b (b - b_o)^2 \]

\[ V_{dihedral} = K_\phi (1 + \cos n\phi - \delta) \]

\[ V_{angle} = K_\theta (\theta - \theta_o)^2 \]

Figure 1.5: Interaction Terms in Potential Energy Function

1.4.3 Docking Simulations

Complexes are formed between molecules that are complementary in structure as well as in physical-chemical properties, and intermolecular van der Waals interactions, electrostatic interactions, and hydrogen bonds stabilize molecular complexes. Docking simulations involve predicting the structure of a protein-ligand or a protein-protein complex when the coordinates of the components are known. Because of the difficulty involved in crystallizing complexes, the
prediction of the structure of a complex through computational means, if done through accurate
docking methods, could be useful in obtaining structural data about complexes, which could be
used as basis for future experimental design or to derive relevant functional information. There
are two important stages in any docking program, first a search of the configurational and
conformational degrees of freedom and then a scoring or evaluation function to rank the
complexes. In the initial stage the components are combined based on geometric, chemical
complementarity and other criteria, then a scoring function is applied to filter the docking results
and a final refinement of the structures is done through molecular mechanics techniques.

Prediction of protein-protein complexes is inherently more difficult than predicting
protein-small molecule complexes due to the limitations in exploring large molecular surfaces
and finding the most appropriate fit.\textsuperscript{41} The success rate and efficiency of docking depends on
several factors like the algorithm used for the docking search problem, the scoring function
applied and the extent of biological information available to filter the results. Since all structure
prediction techniques depend one way or another on experimental data, the availability of sound
background information gives more confidence to the structure prediction and modeling results.

1.5 Overview of Chapters

1.5.1 Quantitative Structure-Activity Relationship Studies of Choline Acetyltransferase
Inhibitors

Chapter 2 explores the structure-activity profile of \textit{trans}-1-methyl-4-(1-naphthylviny1)
pyridinium (MNVP\textsuperscript{+}) analogs to predict structural features that are essential for selective and
potent choline acetyltransferase (ChAT) inhibition. Although ChAT is a well known enzyme,
important for the biosynthesis of acetylcholine (ACh), there is limited structural information
available. In this study, three-dimensional quantitative structure-activity relationship (3D-QSAR)
analyses of some known ChAT inhibitors reveal a pharmacophore model that establishes the optimum molecular features required for the inhibitory activity. The study also provides information regarding the ligand-binding site of this neurologically important ChAT enzyme receptor, which could be used to predict the activity of new molecules prior to synthesis.\textsuperscript{18,19}

Organophosphorus nerve gases such as sarin, soman and VX act mainly by inhibiting acetylcholinesterase (AChE) at cholinergic synapses. The consequent accumulation of high doses of ACh causes seizures and death from respiratory failure.\textsuperscript{42} There has been some interesting speculation that ChAT inhibitors may be useful for treating nerve gas poisoning, as the inhibition of ChAT would result in the decreased synthesis of ACh.\textsuperscript{43,44} The idea being that inhibiting the synthesis of ACh could counter the elevation of Ach levels caused by nerve gas agents and thereby ameliorate its excessive action. Also, it was observed that ChAT inhibitors were able to slow the rate of aging of nerve gas-inhibited AChE and this could be helpful since this would provide more time for reactivating agents to activate AChE. This approach is also better than using muscarinic antagonists like atropine to quell ACh action, as nerve gas agents target both types of cholinergic receptor - muscarinic and nicotinic, whereas atropine primarily acts on muscarinic receptors.\textsuperscript{45-48}

The application of two 3D-QSAR techniques, Comparative Molecular Field Analysis (CoMFA)\textsuperscript{20,21} and Comparative Molecular Similarity Indices Analysis (CoMSIA)\textsuperscript{22,23} to the set of 21 MNVP\textsuperscript{+} analogs are described. These methods have been used to gain insight into the steric, the electrostatic, the hydrophobic, the hydrogen-bonding properties of these molecules, their influence on the activity and to derive predictive 3D-QSAR models for design and prediction of the activities of new analogs for this class of inhibitors. After the 3-D QSAR models were established for our structures, graphical examination of the model depicted which
areas around the molecules should contain a greater or lesser amount of electronegative and
electropositive groups and whether steric factors were an important factor in the inhibition of ChAT.
Apart from CoMFA and CoMSIA descriptor variables, Moriguchi LogP values for the
compounds were included in the development of the models, which complement the field
descriptors by correlating with biotransport phenomena rather than specific receptor
interactions. This improved the results markedly by giving a higher correlation coefficient.
The CoMSIA model performed better than the CoMFA model in terms of its predictive abilities.
In general, the 3D-QSAR approach gives an insight into the different field contributions around
the molecules, and their influence on the overall activity.

1.5.2 Computational modeling of Nuclear Factor κ-B (NF-κB) p50 homodimer

The evolution of HIV infection and the progression of immunosuppression, which is associated
with an increased activation of latent virus, are influenced by various factors. These factors can
induce HIV expression by acting at the transcriptional and post transcriptional level and all the
factors have one approach in common - to create oxidative stress involving reactive oxygen
species (ROS), which leads to the activation of transcription factors including the nuclear
transcription factor (NF-κB), which then further induces HIV expression with the disease
progressing into AIDS.

In Chapter 3 the oxidized form of the transcription factor NF-κB is investigated through
computational modeling and molecular dynamics methods to study the nuclear translocation of
the homodimer. The molecular modeling results help to determine how the disulfide bridge
formation between the two subunits of the dimer facilitates nuclear entry, through comparison
with the experimentally known structure of the reduced form.51
NF-κB belongs to a family of inducible mammalian transcription factors, the Rel/NF-κB family. The Rel homology domain (RHD) contains conserved amino acid sequences in all NF-κB/Rel proteins, involved in important functions including DNA binding, dimerization, nuclear localization, and interaction with IκB molecules.52 Many homo and hetero-dimeric forms of NF-κB have been described and among these the NF-κB heterodimer consisting of two subunit molecules, p65 and p50 is the most well known. It usually exists as a molecular complex with an inhibitory molecule, IκB, in the cytosol.53-56 The function of NF-κB is regulated by the intracellular reduction-oxidation (redox) state.57-61 The initial activation of NF-κB in the cytosol is stimulated by oxidative processes like exposure to reactive ROS. The oxidative stimuli induce a sequence specific phosphorylation and ubiquitination of the I-κB molecule from the NF-κB-I-κB complex, and positively charged nuclear location sequences (NLS) in the two subunits are unmasked.62 This results in the exposure of nuclear localization signal (NLS) in NF-κB and translocates it into the nucleus where it can bind to its cognate DNA binding sites to activate gene transcription.63-66

It is hypothesized that after the removal of I-κB, the two subunits of the NF-κB homodimer adopt a closed conformation through a hinge movement involving the flexible linker connecting the amino-terminal and dimerization domains of p50. This conformational change makes the dimer structure compact and facilitates its nuclear translocation. Starting from the crystal structure of NF-κB p50 homodimer, the peptide region that confers flexibility to the dimer was identified and the backbone torsion angles were systematically modified to establish optimal structural parameters. The localized torsion angle modification induced a large movement about the hinge thereby bringing the cysteine residues (Cys 62) of the two subunits closer to each other for disulfide bonding. Molecular dynamics simulation yielded a refined
structure which showed that the oxidized form of NF-κB was energetically more favored than the reduced form and is significantly more compact than the reduced form which would enable its easier entry through the nuclear pores. Although it has been established that thioredoxin (Trx) regulates the DNA binding activity of NF-κB by reduction of a disulfide bond involving Cys 62 residues, the exact structure of NF-κB in its oxidized form has not been determined. The model derived from this simulation method describes the structure of NF-κB in its oxidized state in the cytosol.

1.5.3 Vascular endothelial growth factor – Molecular docking and analysis of interactions

In Chapter 4 the structure of protein-protein complex between vascular endothelial growth factor (VEGF) with anti-angiogenic molecule, SPARC (Secreted Protein Acidic and Rich in Cysteine) is predicted through molecular docking methods using shape and electrostatic complementarity in an effort to understand the specific interactions. Molecular dynamics methods were used to refine the structural models following which side-chain interactions were identified at the protein interface that are predicted to contribute a large fraction of binding free energy. These identified interactions will be used for directing further mutagenesis studies to investigate their effect on the binding activity. The docking model obtained provides us a detailed understanding of the protein-protein interactions and can be used as a basis for guiding future studies aimed at identifying inhibitors for VEGF-induced angiogenesis.

Angiogenesis, the growth of new blood vessels from pre-existing ones is a key event in both physiological and pathological events. It is a complex biological process that requires the precise coordination of multiple different steps. All of them are regulated by a delicate balance of a variety of angiogenic and angiostatic factors. VEGF is the most important and well characterized angiogenic molecule which is involved in the regulation of multiple steps of
angiogenesis. VEGF binds to two receptors: VEGFR-1 (flt-1) and VEGFR-2 (KDR) which are composed of 7 immunoglobulin domain in the extracellular part and an intracellular tyrosine kinase domain.  

SPARC is an angiostatic factor known to inhibit cellular proliferation and also regulate the activity of certain growth factors such as platelet-derived growth factor (PDGF) and VEGF. Recent studies have shown that SPARC binds to VEGF and inhibits VEGF-stimulated proliferation of endothelial cells. It is hypothesized that this binding site could overlap at least partially the binding site of VEGFR-1, as SPARC acts by preventing VEGF-induced phosphorylation of VEGFR-1.

In order to understand the basis for the multifunctional role of SPARC, further knowledge regarding its interactions with VEGF is needed. To this end the protein-protein docking simulations were carried out using FTDock to understand the interactions between these molecules and their impact on angiogenesis. Mutagenesis data was used to filter the docking results and the residues implicated by mutagenesis experiments as having an influence on the interaction, were used to guide the manual selection of docked complexes. The docking model obtained provides us with a detailed understanding of the protein-protein interactions and can be used as a basis for guiding future studies aimed at identifying inhibitors for VEGF-induced angiogenesis.

1.6 References

5. Trend Watch: Success Rate by Discovery Phase, New and Established Target Projects, Drug Discovery & Development 2005.


40. L. Pauling, M. Delbruck, The nature of the intermolecular forces operative in biological processes, Science 92 (1940) 77–79.


57. R. Brigelius-Flohe, Interleukin-1-induced nuclear factor kappa B activation is inhibited by overexpression of phospholipid hydroperoxide glutathione peroxidase in a human endothelial cell line, Biochemical Journal 328 (1997) 199-203.


CHAPTER 2

THREE DIMENSIONAL QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP
(3D-QSAR) ANALYSES OF CHOLINE ACETYLTRANSFERASE INHIBITORS

Reprinted here with permission of publisher.
2.1 Abstract

As a basis for predicting structural features that may lead to the design of more potent and selective inhibitors of Choline Acetyltransferase (ChAT), the three-dimensional quantitative structure-activity relationship (3D-QSAR) studies were carried out on a series of trans-1-methyl-4-(1-naphthylvinyl)pyridinium (MNVP⁺) analogs, which are known ChAT inhibitors. 3D-QSAR studies were carried out using the Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) methods. Since these inhibitors have extremely shallow potential energy minimum energy wells and low barriers to rotation, two dihedral angles unique to these inhibitors were systematically modified to reflect the energetically preferred conformations as determined by force field calculations. An optimum alignment rule was devised based on the conformations obtained from the molecular mechanics studies, using a common substructure alignment method. The studies involve a set of 21 compounds and experimentally determined molar IC₅₀ values were used as the dependent variable in the analysis. The 3D-QSAR models have conventional r² values of 0.953 and 0.954 for CoMFA and CoMSIA respectively; similarly, cross-validated coefficient q² values of 0.755 and 0.834 for CoMFA and CoMSIA, respectively, were obtained. On the basis of these predictive r² values the model was tested using previously determined IC₅₀ values. CoMSIA 3D-QSAR yielded better results than CoMFA.
2.2 Introduction

Choline Acetyltransferase (ChAT) is an important enzyme in the biosynthesis of acetylcholine. Although the enzyme is well known, there is limited structural information available. Over the years, attempts have been made to gain insights into the ChAT binding cavity through indirect means, primarily by probing the interaction of small molecule inhibitor analogs of trans-1-methyl-4-(1-naphthylvinyl)pyridinium (MNVP⁺) and related compounds. Although the direct inhibition of ChAT does not appear to be a viable therapeutic strategy for various disease states at this time, there has been some interesting speculation that ChAT inhibitors may be useful for nerve gas poisoning. The experimental data based on mouse model studies, seem to suggest that the mechanism may not be directly correlated to ChAT enzyme inhibition. Another intriguing approach of potential therapeutic benefit concerns the use ChAT stimulants to increase the levels of acetylcholine. Regardless of the eventual outcome for potential therapeutic approaches involving ChAT inhibition or stimulation, it is of fundamental importance to have a better understanding of this enzyme.

Significant investments and resources have been made by those involved in computational chemistry and computer-assisted drug design to design compounds that will interact with a particular receptor site when the three-dimensional data available for that receptor site is sparse or lacking altogether. Unfortunately, this is the situation that is often found in pharmaceutical research and one medicinal chemists must face. The absence of detailed structural information about the receptor binding site, however, does not preclude the possibility of inferring a pharmacophore model based on only the ligand structure. Three-dimensional quantitative structure activity relationship (3D-QSAR) methods facilitate the correlation of the three-dimensional structures of potential drug molecules with their biological activity and in
These methods are based entirely on experimental structure-activity relationships for receptor ligands or enzyme inhibitors, and their application in the last 30 years led to several drugs currently on the market.\textsuperscript{3,4,5}

Inhibitors of Choline Acetyltransferase (ChAT) have been known since the late 1960’s. Analogs of \textit{trans}-1-methyl-4-(1-naphthylvinyl)pyridinium (MNVP\textsuperscript{+}) (Figure 2.1)\textsuperscript{6-12} are the simplest and most potent prototype. There are a few known features of MNVP\textsuperscript{+} which appear essential for ChAT inhibition. An essentially coplanar conformation of the \textit{trans} a-b-c region is believed to be important, although it does not cost much energy to adopt twisted conformations. The feasibility of the co-planarity hypothesis has been supported in an earlier computational chemistry study.\textsuperscript{10} Optimum molecular features governing inhibitory potency include region a as a bicyclic aryl group or as a phenyl moiety with \pi-electron donor enhancing substituents, region b as a site of unsaturation typically with a double bond (a triple bond, although not used in this study, is also seen) and region c as a pyridinium or quinolinium system. Substituents in region d are the least structurally refined at this time. The quaternary “onium” forms of these ChAT inhibitors pose limitations in whole animal studies since such ionized species pass through biological membranes with difficulty.

Although ring systems other than pyridinium and quinolinium have been examined (c portion of Figure 2.1), the MNVP\textsuperscript{+} analogs comprise the largest family of molecular variations of ChAT inhibitors and include members for which activities have been confirmed by various researchers.\textsuperscript{11} Despite their pharmacological limitations, these compounds still may be useful as biochemical probes to gain further insight into the three-dimensional structure of the neurologically important ChAT enzyme receptor. To provide the highest possible congruency
among the compounds included in this first 3D-QSAR study of MNVP$^+$ analogs, all of the compounds examined had been assayed for their inhibitory potency by the same procedure.$^{8,9,12}$

\[ \text{Figure 2.1: trans-1-methyl-4-(1-naphthylvinyl)pyridinium} \]

In this paper, the application of two three-dimensional quantitative structure activity relationship (3D-QSAR) methods, Comparative Molecular Field Analysis (CoMFA)$^{13}$ and Comparative Molecular Similarity Indices Analysis (CoMSIA)$^{14,15}$ for MNVP$^+$ analogs, as ChAT inhibitors are described. These ligand-based methods of analysis are used widely since they are not very computationally intensive and afford rapid generation of QSAR’s from which biological activity of newly designed compounds can be predicted.$^3$ In order to make use of CoMFA and CoMSIA, the following four procedures are required: (1) superposition of a set of molecules whose activities have been measured; (2) computation of interaction energy fields with various probes; (3) statistical analyses to correlate the fields with activities and; (4) interpretation of the coefficients of the resulting QSAR equations.$^{16}$

Based on the alignment of the molecules and the interaction of the steric and electrostatic variables of each molecule with a defined probe atom residing at lattice points in a region in space surrounding the molecules, energy contributions to the interaction between the compounds
and the amino acids in the form of lattice points in the active site are sampled. After this region is generated, the results are compared to the pharmacological data, and a linear combination of these two sets of data is constructed using a Partial Least Squares (PLS) algorithm. Cross-validated and non cross-validated $r^2$ values are determined based on the PLS results in order to validate the predictive properties of the model. The $r^2$ values can be optimized by iteratively varying the alignment rules, conformations and other parameters inherent to the technique. This general procedure has been used in the present study to gain insight into the steric, the electrostatic, the hydrophobic, the hydrogen-bonding properties of these molecules, their influence on the activity and to derive predictive 3D-QSAR models for design and prediction of the activities of new analogs for this class of inhibitors.

2.3 Materials and Methods

2.3.1 Computational details

The molecular modeling software SYBYL 6.9\textsuperscript{17} installed on a Silicon Graphics workstation with IRIX 6.5 operating system, was used for three-dimensional structure generation and molecular modeling studies. The force field calculations used MM3 (2000) with parameters developed for positively charged nitrogen atom and the charges were calculated using the Gasteiger-Huckel method as implemented in SYBYL. The low energy conformers and relative energies for all ChAT inhibitors were examined through dihedral driver optimization technique. The minimum structures were examined, constraints were relaxed and frequency calculations were carried out to ensure that a global minimum was achieved.

2.3.2 Conformational Searching

In the present study, since the structural information on these inhibitor-protein complexes are not available; the conformation of the molecules was obtained from systematic
conformational search procedures. The torsion angles $\alpha$ and $\beta$ were systematically driven in 15° increments, using the dihedral driver option available in SYBYL (Figure 2.1). The low energy conformers were examined and the frequency calculations ensured that the global minimum had been identified on the potential energy surface. It was determined that a low barrier to rotation of both $\alpha$ and $\beta$ torsions existed. The torsion angles which reflect the minimum structure demonstrate that all of the compounds in this study exhibited essentially coplanar conformations. These structural observations are consistent with known experimental data in which the trans configuration is favored over cis configuration. Thus, the coplanar conformations were used for the final 3D-QSAR examination of the compounds.

2.3.3 Structural Alignment

The most important requirement for 3D-QSAR techniques (CoMFA and CoMSIA) is that the 3D structures of the molecules to be analyzed be aligned according to a suitable conformational template, which is assumed to adopt a “bioactive conformation”. The molecules in the database were aligned using the “Database Align” routine available in SYBYL. The compounds were fitted to the template molecule 1, one of the most active molecules. The alignment rule was optimized by varying the common substructure used for alignment. As an initial attempt, the structures in the database were aligned by overlaying the carbons in the vinyl linkage. This was thought to be a reasonable approach since this region held the two aromatic moieties together. Another attempt tried to overlay the structures in the database using the pyridinium or quinolinium ring as the backbone. This approach, however, proved to be unproductive. The most successful alignment rule overlaid the atoms of the phenyl moiety in the region a (Figure 2.1). Such an alignment placed less significance on the region c and
correspondingly more significance on the region \( a \) of the MNVP\(^+\) analogs. The entire database was aligned using the coplanar conformations as defined previously and is shown in Figure 2.2.

### 2.3.4 Biological data

The activity data for all the 21 analogs were obtained from the literature reported by Cavallito et al.\(^8,12\). All of the compounds examined were assayed for their inhibitory potency by the same procedure, to avoid any incongruency of data. The activity data for each compound is listed in Table 2.1. For this study, the negative log of molar IC\(_{50}\) (-log IC\(_{50}\)) values reported were used as it would give numerically larger data values for the active compounds than those of the inactive compounds. This is merely a convenience as most of the analysis programs are designed to look for a maximum response. Moriguchi LogP values were also calculated for each of the compounds in the data set.\(^{19,20}\)

![Figure 2.2: Alignment of the ChAT inhibitors using common substructure method.](image-url)
Table 2.1: Cholinergic activity and Moriguchi LogP values used in the CoMFA and CoMSIA study

![Diagram](image)

<table>
<thead>
<tr>
<th>Compound</th>
<th>-log [molar IC&lt;sub&gt;50&lt;/sub&gt;]</th>
<th>logP&lt;sup&gt;l&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-Naphthyl</td>
<td>6.33</td>
</tr>
<tr>
<td>2</td>
<td>3-BrC&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>5.83</td>
</tr>
<tr>
<td>3</td>
<td>3-ClC&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>5.64</td>
</tr>
<tr>
<td>4</td>
<td>4-BrC&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>5.59</td>
</tr>
<tr>
<td>5</td>
<td>3,4-Cl&lt;sub&gt;2&lt;/sub&gt;C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;3&lt;/sub&gt;</td>
<td>5.48</td>
</tr>
<tr>
<td>6</td>
<td>4-ClC&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>5.16</td>
</tr>
<tr>
<td>7</td>
<td>Phenyl</td>
<td>4.83</td>
</tr>
<tr>
<td>8</td>
<td>4-CH&lt;sub&gt;3&lt;/sub&gt;C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>4.70</td>
</tr>
<tr>
<td>9</td>
<td>4-FC&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>4.68</td>
</tr>
<tr>
<td>10</td>
<td>2-ClC&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>4.56</td>
</tr>
<tr>
<td>11</td>
<td>3-NO&lt;sub&gt;2&lt;/sub&gt;C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>3.83</td>
</tr>
<tr>
<td>12</td>
<td>4-NO&lt;sub&gt;2&lt;/sub&gt;C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>3.70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>-log [molar IC&lt;sub&gt;50&lt;/sub&gt;]</th>
<th>logP&lt;sup&gt;l&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>1-Naphthyl</td>
<td>6.40</td>
</tr>
<tr>
<td>14</td>
<td>3-ClC&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>5.80</td>
</tr>
<tr>
<td>15</td>
<td>3,4-Cl&lt;sub&gt;2&lt;/sub&gt;C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;3&lt;/sub&gt;</td>
<td>5.77</td>
</tr>
<tr>
<td>16</td>
<td>Phenyl</td>
<td>5.30</td>
</tr>
<tr>
<td>17</td>
<td>2-ClC&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>4.77</td>
</tr>
<tr>
<td>18</td>
<td>3-Indolyl</td>
<td>4.41</td>
</tr>
</tbody>
</table>
2.3.5 CoMFA and CoMSIA 3D-QSAR models

CoMFA and CoMSIA methods were performed with the QSAR option of SYBYL. For all steps, the default Sybyl settings were used except otherwise noted. In deriving the CoMFA and CoMSIA descriptor fields a 3D cubic lattice with a grid spacing of 2.0 Å in x, y and z directions was created to encompass the aligned molecules.

The CoMFA steric (Lennard-Jones 6-12 potential) and electrostatic (Coulomb potential) field energies were calculated using sp³ carbon probe atom carrying +1 charge, with a distance dependent dielectric at each lattice point. The charges were determined using the Gasteiger-Huckel method. The energy calculation was performed for all grid points such that all energies were constrained to be between -30 and +30 kcal/mol.

CoMSIA descriptors were also derived using the standard implementation in the Sybyl package. Five physicochemical properties (steric, electrostatic, hydrophobic²¹, hydrogen bond donor and hydrogen bond acceptor) were evaluated using the probe atom. The probe atom used has a radius of 1Å, charge of +1, hydrophobicity of +1, and hydrogen bonding and hydrogen

<table>
<thead>
<tr>
<th>Compound</th>
<th>-log [molar IC₅₀]</th>
<th>logP</th>
</tr>
</thead>
<tbody>
<tr>
<td>19 Phenyl</td>
<td>5.46</td>
<td>4.15</td>
</tr>
<tr>
<td>20 2-Thienyl</td>
<td>3.60</td>
<td>3.35</td>
</tr>
<tr>
<td>21 3-Pyridyl</td>
<td>3.22</td>
<td>2.64</td>
</tr>
</tbody>
</table>
donor properties of +1. These fields were selected to cover the major contributions to ligand binding.\textsuperscript{22} The default value of 0.3 was used as the attenuation factor ($\alpha$).\textsuperscript{15}

A Partial least Squares (PLS) analysis was done to derive the 3D-QSAR models, utilizing the CoMFA standard scaling for the molecular fields. The activity data, $-\log IC_{50}$, was used as the dependent variable and the predictive value of the model, represented by $q^2$, was evaluated using the leave-one-out (LOO) cross validation method. The optimum number of components was determined based on the standard error of prediction and $q^2$ values. The equation that gave a low standard error of prediction and high $q^2$ was chosen. The conventional correlation coefficient $r^2$ values were also calculated and the incorporation of Moriguchi LogP values improved the PLS results markedly.

After the 3-D QSAR models were established for our structures, graphical examination of the model depicted which areas around the molecules should contain a greater or lesser amount of electronegative and electropositive groups and whether steric was an important factor in the inhibition of ChAT. The “Predict Property” command in Sybyl was utilized to determine what the predicted potencies of these compounds were based on the model developed using CoMFA and CoMSIA. The actual and the predicted values of the compounds are listed in Table 2.3.

2.4 Results and Discussion

The 3D-QSAR models which were derived, employing the CoMFA and CoMSIA methods consisting of a set of 21 compounds, (Table 2.1) must be consistent with the spatial, electronic and other variables that act in concert to provide optimum inhibitory activity. Several of these variables have been identified or proposed for the a-b-c-d components of the molecule and for the molecule as a whole.\textsuperscript{10} Component b which is a $trans$ vinyl linkage and component d
which primarily serves to quaternize the N atom, do not appear to be structurally specific. This leaves the burden of establishing optimum structural parameters on a and c.

Apart from CoMFA and CoMSIA descriptor variables, Moriguchi LogP values for the compounds were included in the development of the models, which complement the field descriptors by correlating with biotransport phenomena rather than specific receptor interactions. This improved the results markedly by giving a higher correlation coefficient. The statistical data obtained from the PLS method for the CoMFA and CoMSIA models is shown in Table 2.2.

CoMFA models were developed using a common substructure alignment method. The cross-validated \( q^2 \) value from the PLS regression analysis is 0.755, while the non cross-validated \( r^2 \) with five components is 0.953. The actual and predicted values of the activities for the compounds are listed in Table 2.3.

The CoMFA steric contour plot obtained is shown in Figure 2.3. The green (sterically favorable) and yellow (sterically unfavorable) regions in the contour represent 80% and 20% level contributions, respectively. A part of the structure of the most active compounds 1 and 13 protrude into the green (sterically favorable) contour region whereas the other relatively less active compounds do not occupy this region (Figure 2.3). The less active compounds either partially touch the sterically unfavorable regions or do not occupy the sterically favorable region (not shown). The slightly lower activity of 12 as compared to 7 may be due to the fact that p-nitro group touches the sterically unfavorable yellow contour region. Thus, the CoMFA steric contours explain the differences in the overall activity of the inhibitors due to differences in spatial arrangement.
Table 2.2: PLS statistical data obtained from the CoMFA and CoMSIA models

<table>
<thead>
<tr>
<th>PLS statistics</th>
<th>CoMFA</th>
<th>CoMSIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q^2_a$</td>
<td>0.755</td>
<td>0.834</td>
</tr>
<tr>
<td>$r^2_b$</td>
<td>0.953</td>
<td>0.954</td>
</tr>
<tr>
<td>ONC$^c$</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>F value$^d$</td>
<td>53.109</td>
<td>53.737</td>
</tr>
<tr>
<td>SEE$^e$</td>
<td>0.197</td>
<td>0.196</td>
</tr>
<tr>
<td>Field Contributions$^f$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LogP</td>
<td>8.3%</td>
<td>22.3%</td>
</tr>
<tr>
<td>Steric</td>
<td>61.3%</td>
<td>9.8%</td>
</tr>
<tr>
<td>Electrostatic</td>
<td>30.4%</td>
<td>18.9%</td>
</tr>
<tr>
<td>Hydrophobic</td>
<td>--</td>
<td>30.7%</td>
</tr>
<tr>
<td>H-bond donor</td>
<td>--</td>
<td>6.8%</td>
</tr>
<tr>
<td>H-bond acceptor</td>
<td>--</td>
<td>11.5%</td>
</tr>
</tbody>
</table>

$^a$Cross-validated correlation coefficient.
$^b$Non cross-validated correlation coefficient.
$^c$Optimal number of components obtained form cross-validated PLS.
$^d$F-Test value.
$^e$Standard error of estimate.
$^f$Field contributions: LogP, steric and electrostatic fields from CoMFA. LogP, steric, electrostatic, hydrophobic, H-bond donor/acceptor fields from CoMSIA.

The CoMFA electrostatic contour is shown in Figure 2.4. The blue (positive charge favored region) and red (negative charge favored region) contours represent 80% and 20% level contribution, respectively. The slightly higher activities of compounds 2 and 3 as compared to 4, 9 and 10, although all these compounds have an electronegative group attached to them, is due to the fact that the halogen group in compounds 2 and 3 occupy the red contour region, which is favorable for electronegative groups while the other compounds do not occupy that region. Overall compounds that show relatively higher activities in the set either have their electronegative group in the red contour or the pyridinium nitrogen near the blue contour.
Table 2.3: Results of the predicted activities and residuals of the compounds by the CoMFA and CoMSIA models

<table>
<thead>
<tr>
<th>Compd</th>
<th>Actual</th>
<th>Predicted (CoMFA)</th>
<th>Residual</th>
<th>Predicted (CoMSIA)</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>6.40</td>
<td>6.378</td>
<td>0.022</td>
<td>6.503</td>
<td>-0.103</td>
</tr>
<tr>
<td>1</td>
<td>6.33</td>
<td>6.343</td>
<td>-0.013</td>
<td>6.318</td>
<td>0.012</td>
</tr>
<tr>
<td>2</td>
<td>5.83</td>
<td>5.753</td>
<td>0.077</td>
<td>5.881</td>
<td>-0.051</td>
</tr>
<tr>
<td>14</td>
<td>5.80</td>
<td>6.018</td>
<td>-0.218</td>
<td>5.782</td>
<td>0.018</td>
</tr>
<tr>
<td>15</td>
<td>5.77</td>
<td>5.681</td>
<td>0.089</td>
<td>5.760</td>
<td>0.010</td>
</tr>
<tr>
<td>3</td>
<td>5.64</td>
<td>5.543</td>
<td>0.097</td>
<td>5.582</td>
<td>0.058</td>
</tr>
<tr>
<td>4</td>
<td>5.59</td>
<td>5.250</td>
<td>0.340</td>
<td>5.275</td>
<td>0.315</td>
</tr>
<tr>
<td>5</td>
<td>5.48</td>
<td>5.503</td>
<td>-0.023</td>
<td>5.563</td>
<td>-0.083</td>
</tr>
<tr>
<td>19</td>
<td>5.46</td>
<td>5.440</td>
<td>0.020</td>
<td>5.516</td>
<td>-0.056</td>
</tr>
<tr>
<td>16</td>
<td>5.30</td>
<td>5.206</td>
<td>0.094</td>
<td>4.875</td>
<td>0.425</td>
</tr>
<tr>
<td>6</td>
<td>5.16</td>
<td>5.099</td>
<td>0.061</td>
<td>5.125</td>
<td>0.035</td>
</tr>
<tr>
<td>7</td>
<td>4.83</td>
<td>4.664</td>
<td>0.166</td>
<td>4.646</td>
<td>0.184</td>
</tr>
<tr>
<td>17</td>
<td>4.77</td>
<td>4.914</td>
<td>-0.144</td>
<td>4.822</td>
<td>-0.052</td>
</tr>
<tr>
<td>8</td>
<td>4.70</td>
<td>4.886</td>
<td>-0.186</td>
<td>4.903</td>
<td>-0.203</td>
</tr>
<tr>
<td>9</td>
<td>4.68</td>
<td>4.928</td>
<td>-0.248</td>
<td>4.920</td>
<td>-0.240</td>
</tr>
<tr>
<td>10</td>
<td>4.56</td>
<td>4.665</td>
<td>-0.105</td>
<td>4.661</td>
<td>-0.101</td>
</tr>
<tr>
<td>18</td>
<td>4.41</td>
<td>4.478</td>
<td>-0.068</td>
<td>4.435</td>
<td>-0.025</td>
</tr>
<tr>
<td>11</td>
<td>3.83</td>
<td>4.036</td>
<td>-0.206</td>
<td>3.766</td>
<td>0.064</td>
</tr>
<tr>
<td>12</td>
<td>3.70</td>
<td>3.456</td>
<td>0.244</td>
<td>3.907</td>
<td>-0.207</td>
</tr>
</tbody>
</table>

Figure 2.3: CoMFA steric S.D. x coefficient contour plot; green contours indicate regions where steric bulk is favorable, whereas yellow contours indicate regions where steric bulk is detrimental.
Component **b** has been well defined where a *trans* vinyl linkage is optimum. No specific binding contribution can be ascribed to the vinyl bridge other than transmission of electrons between the two rings and facilitation of co-planarity of the inhibitor. It was proposed in an earlier study that the double bond is polarized by a mesomeric interaction with the phenyl ring and the pyridyl ring, causing a partial positive charge on the carbon atom adjacent to the benzene ring and a partial negative charge on the carbon atom adjacent to the pyridyl ring. So, this would enable a nucleophilic residue on the enzyme surface to have a strong interaction with the partial positive charge on the $\beta$-carbon to the pyridyl ring.$^{24}$ This is evident from the electrostatic contour shown in Figure 2.4, where the $\beta$-carbon to the pyridyl ring is near the blue contour, consistent with the positive charge.

**Figure 2.4:** CoMFA electrostatic S.D. $x$ coefficient contour plot; blue contours indicate regions where electropositive groups increase activity, whereas red contours indicate regions where electronegative groups increase activity.

CoMSIA is similar to CoMFA, but uses a Gaussian function rather than Coulombic and Lennard-Jones potentials to assess the contribution from different fields. Furthermore, in addition to the steric and electrostatic fields of CoMFA, CoMSIA defines explicit hydrophobic and hydrogen bond donor/acceptor fields, which are not available with standard CoMFA. The
CoMSIA results gave a cross-validated $q^2$ value of 0.834, while the non cross-validated $r^2$ with 5 components was 0.954. The actual and predicted values of activities for the compounds are listed in Table 2.3. The CoMSIA steric and electrostatic plots are similar to those obtained from CoMFA. The additional hydrophobic, hydrogen bond donor/acceptor plots are shown in Figures 2.5-2.7 respectively.

The hydrophobic fields (yellow, hydrophobic group favored; white, hydrophobic group disfavored) show that the naphthyl group in the active compounds 13 and 1 fit well into the yellow region which is favorable for hydrophobic groups. While most of the other less active compounds either partially touch the yellow contour or occupy unfavorable white region.

![Figure 2.5: CoMSIA hydrophobic S.D. x coefficient contour plot; yellow contours indicate regions where hydrophobic groups increase activity, white contours indicate regions where hydrophobic groups decrease activity](image-url)
Figure 2.6: CoMSIA H-bond donor S.D. x coefficient contour plot; cyan contours indicate regions where H-bond donor group increases activity, whereas orange contours indicate regions where H-bond donor group decreases activity.

Figure 2.7: CoMSIA H-bond acceptor S.D. x coefficient contour plot; green contours indicate regions where H-bond acceptor increases activity, whereas red contours indicate regions where H-bond acceptor decreases activity.

With regards to a, earlier studies have proposed that substituents enhancing π-donor qualities are favorable for activity.\textsuperscript{10} This is clearly seen from the H-bond donor contours (cyan, favored; orange, disfavored) where the aryl group in the region a is close to the cyan contour
which is favored for activity. The extended naphthyl group at a increases the \( \pi \)-donor abilities which could assist in hydrogen bonding to a nearby histidine residue.\(^{25} \) Thus, overall the CoMSIA fields depict the areas around the molecules where changes either increase or decrease activity. Figure 2.8 represents graph of the actual versus predicted \(-\log IC_{50}\) of the molecules for CoMFA and CoMSIA 3D QSAR models.

(a)

(b)

**Figure 2.8:** Comparison of actual vs. predicted \(-\log IC_{50}\) of the molecules for (a) CoMFA and (b) CoMSIA 3D QSAR models.
The 3D-QSAR models developed were assessed for their predictive abilities using a test set of 4 compounds, which were not included in the development of the models (Table 2.4).

**Table 2.4:** Chat inhibitory activity and Moriguchi LogP values of the test set molecules

<table>
<thead>
<tr>
<th>Compound</th>
<th>-log [molar IC$_{50}$]</th>
<th>logP$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>22 3-BrC$_6$H$_4$</td>
<td>5.38</td>
<td>3.18</td>
</tr>
<tr>
<td>23 3-ClC$_6$H$_4$</td>
<td>5.62</td>
<td>3.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>-log [molar IC$_{50}$]</th>
<th>logP$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 1-Naphthyl</td>
<td>6.28</td>
<td>4.17</td>
</tr>
<tr>
<td>25 Phenyl</td>
<td>5.54</td>
<td>3.50</td>
</tr>
</tbody>
</table>

The CoMSIA model performed better in predicting the activities than the CoMFA model, possibly because of the inclusion of additional fields to correlate with the activity. The advantage
of the ethanol group in these compounds provides adequate aqueous solubility for testing. Variations at this end, especially with extended alkyl groups, may aid in increasing the lipophilicity of these inhibitors. Also, it is conceivable that the alignment of the pyridinium system is not involved in ChAT binding at all. The actual and predicted activities of these compounds are shown in Table 2.5.

<table>
<thead>
<tr>
<th>Compd</th>
<th>Actual</th>
<th>Predicted (CoMSIA)</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>6.28</td>
<td>5.69</td>
<td>0.59</td>
</tr>
<tr>
<td>23</td>
<td>5.62</td>
<td>5.28</td>
<td>0.34</td>
</tr>
<tr>
<td>25</td>
<td>5.54</td>
<td>5.66</td>
<td>-0.12</td>
</tr>
<tr>
<td>22</td>
<td>5.38</td>
<td>5.69</td>
<td>-0.31</td>
</tr>
</tbody>
</table>

2.5 Conclusion:

We have developed predictive CoMFA and CoMSIA 3D-QSAR models for MNVP⁺ analogs as ChAT inhibitors. The CoMSIA model performs better than the CoMFA model in terms of its predictive abilities. These proposed models can be used to predict the activity of newly designed analogs, prior to synthesis. The contour diagrams can be used to identify the structural features of new compounds that would increase or decrease their potency and their activities can be predicted using the QSAR equations. In general, the 3D-QSAR approach gives an insight into the different field contributions around the molecules, and their influence on the overall activity. Further refinement of the structural features might lead to increased potency and aid in increasing the lipophilicity of the molecules. This idea is under investigation in our laboratory.
2.6 Acknowledgement

This work was carried out at the University of Georgia.

2.7 References


17. The program SYBYL 6.9 is available from Tripos Inc., 1669, South Hanley Road, St.Louis, MO 63144, USA.


CHAPTER 3

MOLECULAR MODELING OF THE OXIDIZED FORM OF NUCLEAR FACTOR-κB SUGGESTS A MECHANISM FOR REDOX REGULATION OF DNA BINDING AND TRANSCRIPTIONAL ACTIVATION

\[1\] Chandrasekaran, V., and E.W. Taylor. To be submitted to *Journal of Molecular Graphics and Modelling.*
3.1 Abstract

NF-κB is an important transcriptional regulator of numerous cellular genes, as well as retroviruses such as HIV-1. Paradoxically, oxidative stimuli in the cytosol are associated with nuclear translocation of NF-κB, but in the nucleus, reductive activation by thioredoxin is required for NF-κB to bind to DNA and activate target genes. Experimental structures of the reduced form of NF-κB and its DNA targets are available, from which we modeled the oxidized form of NF-κB homodimer by removal of bound DNA, and modification via a hinge movement of a linker between the dimerization and DNA binding domains of each subunit. Molecular dynamics then enabled the formation of an inter-subunit disulfide bond between the Cys59 residues of each monomer. The resulting model of oxidized, disulfide-bridged NF-κB is clearly more compact than the open, reduced form, which may explain why oxidation is necessary for nuclear translocation, through pores in the nuclear envelope. Furthermore, the inter-subunit disulfide blocks DNA from entering the active site of the oxidized dimer, explaining why reduction to the thiol form in the nucleus is essential for transcriptional activation.
3.2 Introduction

Nuclear factor-κB (NF-κB) is a latent, primary cellular transcription factor that regulates the expression of a wide variety of cellular genes, particularly those involved in host defense.\textsuperscript{1-3} It plays a role not only as an evolutionarily conserved coordinating element in an organism’s response to situations of infection, stress and injury, but also as an activator of a number of viral genes including those of HIV.\textsuperscript{4} While NF-κB acts as the central mediator of the immune response in many cells, multiple families of viruses promote their replication, prevent virus-induced apoptosis and mediate the immune response to the invading pathogen by activating NF-κB.\textsuperscript{5,6} Thus the target genes under the regulation of NF-κB include a variety of cellular as well as viral genes.

NF-κB is composed of homo- or heterodimers of different subunits, which are members of a family of structurally related proteins (Rel/NF-κB proteins). Five different Rel proteins have been identified so far: p50, p52, p65, RelB, and c-Rel. Members of the Rel/NF-κB protein family are characterized by the presence of a Rel Homology Domain (RHD), which contains a nuclear localization sequence (NLS) and is involved in sequence-specific DNA binding, dimerization and interaction with inhibitory I-κB proteins.\textsuperscript{7,8} NF-κB predominantly exists as a heterodimer of p50/p65 subunits in the cytoplasm, although p50 homodimers are also commonly observed. Each subunit of the dimer consists of two major domains, the DNA-binding domain towards the amino-terminal responsible for base-specific contacts, and the dimerization and I-kappa B binding domain. A flexible short (10 amino acids) linker connects these two immunoglobulin-like domains and the carboxy-terminal contains the nuclear localization sequence, which is a set of basic amino acid residues, responsible for targeting the molecule into the nucleus.\textsuperscript{9,10}
In unstimulated cells, NF-κB dimers are retained in the cytosol in an inactive form as a consequence of their association with members of another family of proteins called I-κB (inhibitors of κB). Activators of NF-κB induce the dissociation of I-κB from the NF-κB-I-κB ternary complex, by which positively charged nuclear location sequences in the two subunits are unmasked. This activation cascade includes at least two independent pathways: a kinase pathway and a redox-signaling pathway. In response to a variety of stimuli including physical and chemical stresses, cytokines, reactive oxygen intermediates (ROI) and ultraviolet light, the latent cytoplasmic NF-κB/I-κB complex is activated by the multisubunit I-κB kinase (IKK) complex. The IKK complex causes proteolytic degradation of I-κB through sequence-specific phosphorylation and ubiquitination of the I-κB molecule.

After dissociation from I-κB, NF-κB undergoes redox regulation by the cellular reducing catalyst, thioredoxin (Trx). Trx is a small endogenous molecule having a typical cxxc active site amino acid sequence Cys-Gly-Pro-Cys, in which the two Cys residues give the molecule the ability to reduce disulfide bonds. Trx plays dual and opposing roles in the regulation of NF-κB. Overexpression of wild-type Trx suppressed NF-κB activation in the cytoplasm by interfering with the signal to I-κB kinases and blocking the degradation of I-κB, while inside the nucleus Trx enhanced NF-κB dependent transcription by promoting its DNA-binding ability. Thus, the transcriptional activation of NF-κB can be considered as two distinct steps with respect to the cellular compartment with Trx playing different roles in the cytoplasm and in the cell nucleus. This indicates that activation and regulation of the NF-κB complex is a redox-sensitive process dependent on the activity of reducing proteins like Trx.

In the cytoplasm, reactive oxygen intermediates mediate the degradation and release of I-κB from NF-κB and in those oxidizing conditions, the reactive SH groups in NF-κB would be
oxidized. The p50 subunit of NF-κB has seven cysteine residues at positions 62, 88, 119, 124, 162, 262 and 273. Among these the DNA-binding loop of NF-κB contains a characteristic sequence motif with a cysteine and three arginine residues and these cysteine residues in the two subunits are susceptible to oxidation. In order for NF-κB to bind to the DNA, it is necessary for the oxidized cysteine residues in the DNA-binding region to be in the reduced state. Studies have shown that the DNA-binding activity of p50 correlated well with the redox states of Cys-62, indicating that the reduced form of Cys-62 is important for the DNA-binding activity. Under conditions of oxidative stress, NF-κB is activated and translocated into the nucleus and in these conditions these critical cysteine residues form disulfide bond, thus impairing the ability of NF-κB to bind to DNA. In order to understand how Trx binds to NF-κB in the nucleus and reduces the disulfide bond and activates its DNA-binding, it is important to look at the structure of the oxidized form of NF-κB.

Although there is extensive knowledge available regarding NF-κB structure and its signaling responses, there is still very limited understanding of the in vivo dynamics of this pathway. For example, even though it is well known that NF-κB needs to be reduced by thioredoxin inside the nucleus in order to bind to DNA, the precise structure of the oxidized form of NF-κB when it enters the nucleus is still not known. The crystal structures available so far are of the reduced form of homo-dimeric and hetero-dimeric NF-κB complexes bound to DNA. To date there is no snapshot available of the oxidized form of NF-κB. These crystal structures as such present a static picture of the transcription factor bound to DNA whereas the structure of the oxidized form will allow us to understand the intermediate structures in the pathway and the molecular mechanisms involved prior to transcriptional activation. Similarly, studies have shown the importance of reactive oxygen species in the degradation of the NF-κB/I-
κB complex, but the roles of reactive oxygen species and the reactive cysteine residues in the nuclear transport of NF-κB are not clearly understood. Since the NF-κB dimers assume different conformations in the cytosol when bound to I-κB versus the free or DNA-bound structures, the intermediate structures will give a clearer picture of the dynamic nature of the transcription factor and will enhance our understanding of the NF-κB pathway in relation to various signaling mechanisms. In this study, computational modeling and molecular dynamics simulation techniques are applied to the existing experimental structural information to determine the structure of the oxidized form of NF-κB.

3.3 Methods

3.3.1 Model Building

The crystal structure of the transcription factor NF-κB p50 homodimer bound to a palindromic kappa B site of the DNA (pdb entry: 1NFK) was used as the starting structure for building the oxidized model. The bound DNA helix was deleted from the dimer and the structure was then examined to identify the peptide region that confers flexibility to the dimer and causes the movement of the two subunits. Analysis of the protein structure showed that the peptide linker connecting the amino and carboxy-terminal domains of NF-κB could be involved in a hinge movement that produces a change in the orientation of the DNA binding domains of the two subunits. The basic premise behind this paper is that when NF-κB enters the nucleus following I-κB degradation, the reactive Cys residues in the DNA-binding region of the two NF-κB subunits would be in the oxidized form, indicating that the oxidized structure would require the cysteines in the DNA-binding domains of the two subunits to be within disulfide bonding distance. This implies that such a large movement of the domains could come only from the flexible linker region connecting the DNA-binding domain and the dimerization domain of the subunits.
Subsequently, in an effort to model the oxidized, disulfide bridged structure of NF-κB, the backbone torsion angles of the linker region were systematically modified, respecting the existing secondary structure, to identify the preferred conformation.

**Figure 3.1: Three-dimensional structure of NF-κB p50 homodimer (PDB code: 1NFK) with the DNA removed.** The DNA-binding domain and the dimerization domain are shown in red and blue respectively; the residues (242-244) in the linker region that contribute to the hinge movement are shown in green.

Since the linker region was relatively unconstrained by packing and was not part of an α-helix or a β-sheet the hinge movement did not impose any restrictions on hydrogen bonding. The only structural constraint was that the torsion angles remain in the allowed regions of the Ramachandran plot. Two torsions, the ψ angle and ω angle around the proline residue (Pro-243) were modified by 29.7° and 11.7° respectively to produce a conformation with minimum expenditure of energy and having the least steric hindrance. This localized torsion angle
modification induced a large movement about the hinge thereby bringing the cysteine residues (Cys-59) of the two subunits closer to each other, for disulfide bonding (amino acid residue numbering in this section is based on PDB file: 1NFK).

Figure 3.2: Superposition of the NF-κB dimers before (green) and after (cyan) hinge movement. Hinge movement through torsion angle modification brings the DNA-binding domains closer, while the dimerization domain is unaffected.

The rest of the protein however rotated as a rigid body, with the axis of rotation passing through the hinge region. The distance between the Cys-59 residues in the two subunits was reduced from 19.908 Å in the starting crystal structure to 3.797 Å after modifying the two torsion angles. Examination of the Ramachandran plot of the backbone angles of the modified structure using PROCHECK\textsuperscript{34} showed that it falls in the commonly accepted regions of phi-psi space (Fig
3.4). After structural refinement (described below), the resulting oxidized structure was also found to be energetically more favorable than the starting structure.

### 3.3.2 Model Refinement

Molecular dynamics simulation was carried out after hinge movement, in order to refine the positions of side chains in the modeled oxidized structure. Energy minimization with a classical force field can be used to remove unfavorable steric interactions and large deviations from ideal geometry, but molecular dynamics simulation is required to improve rotamer distributions. To pre-condition the modeled structure, energy minimization using the Kollman all-atom force field\(^{35}\) implemented in SYBYL\(^{36}\) was performed for 2000 steps of steepest descent and 3000 steps of conjugate gradient minimization. A distance dependent dielectric function was used with the dielectric constant set to 1 and the non-bonded cutoff was set to 8 Å. The resulting structure was used as the starting structure for the MD simulation. All MD simulations were performed with the AMBER 7.0 molecular simulation package.\(^{37}\) A complete molecular dynamics simulation study of the NF-κB dimer requires simulation performed when the protein is adequately hydrated. However, due to the large size of the system after adding water molecules, it was decided to simplify the molecular model of the protein so that computing resources can be conserved. Therefore the simulation was carried out in vacuo using the BELLY option and the atoms of the residues within 30 Å from the two cysteine residues were allowed motional freedom during the course of the MD trajectory, while the residues outside the radius were restrained. At physiological pH, the starting structure was positively charged, thus in order to make the simulation system electrically neutral, counter ions (Cl\(^-\)) were added using the LEAP module. As a preparatory step for the simulation a 1000 step minimization was run using a combination of steepest descent and conjugate gradient method to relax the chloride ions and to remove the
strong initial forces due to collisions generated by the ions insertion process. A 3 ns (nanosecond) simulation was performed under constant pressure conditions using the parm99 force field to describe the interactions between the protein atoms. As in the minimization, the non-bonded cutoff for electrostatic and Van der Waals was set to 8 Å and a distance dependent dielectric function with the dielectric constant set to 1.0 was used. The simulation was carried out using a 2 fs time step for a total simulation time of 3 ns. The energy profile was then analyzed to compare the structures of the reduced and oxidized models.

3.4 Results and Discussion

It has been established that Trx regulates the DNA binding activity of NF-κB through reduction of a disulfide bond involving Cys-62 residues, however the exact structure of NF-κB in its oxidized form has not been determined so far. In this study we have derived a model using computational modeling and simulation methods that describes the structure of NF-κB in its oxidized state when it enters the cell nucleus. It is known that the three-dimensional structure of NF-κB is typically composed of a homo-dimeric or hetero-dimeric complex of p50 or p50/p65 subunits and each subunit of the complex is comprised of two domains, the amino-terminal DNA-binding domain consisting of 180 amino acids and the carboxy-terminal dimerization domain with approximately 100 amino acid residues. These two domains are connected by a peptide region composed of 10 amino acid residues. The structural data from our studies indicate that this connecting strand forms a hinge between the two domains and provides conformational flexibility to the molecule.

3.4.1 Effect of torsion angle modifications

Large relative domain movements have been observed in many protein structures which provide examples of their flexibility. In the case of NF-κB, earlier studies have shown that the p65
subunit is remarkably flexible and adopts a different conformation upon binding to I-κBα. Compared to its DNA-bound open conformation, the I-κBα-bound p65 amino-terminal domain rotates almost 180° and translates 38 Å towards its carboxy-terminal domain. This movement is also entirely due to the flexible ten amino acid linker connecting the DNA-binding and dimerization domains.

In order to find the exact residues involved in the hinge movement that results in the oxidized form and to establish their optimum structural parameters, the linker region of NF-κB was examined. A minimum number of residues are clearly needed to afford conformational flexibility, however the energetic costs associated with the structure formation set a limit to this number, above which the expenditure of energy would become too costly to sustain the structure. To be most effective, any reasonable mechanism of conformational change would involve the least expenditure of energy, and will therefore occur through minimum torsional angle modifications to the backbone. Keeping this in mind, systematic search of torsional angle space in the linker region pointed to residues 242-244 that produced significant conformational change to the protein structure upon small torsional modifications. At the same time, these torsional changes kept the secondary structural features intact. Relatively small change to the backbone torsion angles yielded a significant change in the conformation of the subunit, causing an en bloc movement of the entire domain. It could be argued that the domain movements would involve a more complex mechanism than simple backbone torsional modification to produce such a large change in conformation. However, this hinge movement is consistent with domain movements observed in several other proteins, indicating that the flexibility of protein structures is important for their catalytic activity.39,40
The large conformational change occurring in the NF-κB structure upon torsional modification can be described as a rigid body movement of one domain in relation to the other with very little change to the conformation of either domain. The root mean square deviations after superposition of the backbone atoms of residues in the dimerization domain of the two structures is 0.16 Å, while it is 5.86 Å for the DNA-binding domain. Thus the large relative movement of the two subunits is a consequence of small rotations of two backbone torsion angles in the connecting strand. The transcription factor undergoes a dynamic change between the open and closed conformation due to the flexible linker region between the DNA-binding and the dimerization domains. This opening-closing movement of the two domains is common and has been observed in many binding proteins and the peptide region connecting the domains is primarily responsible for the conformational change.\textsuperscript{41} Despite the relatively large number of crystal structures available for the bound conformations of NF-κB with the DNA or I-κB, there
were no structures available for the unbound conformation. In this study, we present the structure of the oxidized form of NF-κB with the Cys-62 residues forming disulfide bond and our data is consistent with previous reports showing that these critical Cys residues need to be reduced by Trx in order for NF-κB to bind to DNA.

3.4.2 Refinement using molecular dynamics simulation

In general, a number of bad steric contacts are introduced during torsion angle modification and in order to eliminate these bad contacts and establish an accurate structural model a MD simulation was carried out. For large proteins, the use of all-atom unrestricted MD simulation to improve models does not produce reliable structure refinement at present, even when the solvent is treated in atomic detail and long-range interactions are considered, due to complications from the size of the system. The sampling of configurational space for a whole protein is too slow to allow sufficient sampling in the available time. But MD simulations are capable of performing better when structure restraints are used to restrict atomic freedom and improved structure refinement results are obtained. Also, the required sampling is less extensive, and subtle insufficiencies in the force fields do not have pronounced effects. In our simulation the primary focus is on the movement of the DNA-binding domain and since NF-κB is a large protein with 624 amino acids, the objective was to simulate a truncated portion of the molecule which would allow for more efficient use of computing time. Therefore, we used a model in which only the atoms within 30 Å distance from the two cysteine residues were included in the simulation. Bad steric contacts were removed and side chain rotamer distributions were improved after a 3 ns simulation. The Ramachandran plot for the oxidized model was computed with PROCHECK to locate any abnormal peptide linkages or other unlikely bond angles. The plot (Figure 3.4) shows
that the residues fall in the allowed regions of phi-psi space and the model is structurally refined after the MD simulation.

**Figure 3.4: Ramachandran plot for the oxidized form of NF-κB**

Energies, both total energy and components of different groups, were calculated and it was observed that the oxidized structure was energetically more favorable than the reduced structure. Comparison of the energy minimized models of the reduced and oxidized forms of NF-κB.
κB shows that the oxidized form is energetically preferred by around 200 Kcals/mol. Since our objective was to test the hypothesis that the oxidized form does not impose any energetic penalty on NF-κB for nuclear transport and is at least as energetically preferable as the reduced form, we carried out the MD simulation to refine any major structural anomalies. The use of MD simulation is limited to refining the structural model and extended simulation times are required for an accurate representation of interactions both within the protein and the environment.

3.4.3 Size of the molecule and ability to bind to DNA

The NF-κB molecules act as substrates for the nuclear transport receptors, importin α3 and α4, and are delivered to the nucleus where it stimulates transcription of genes. Following I-κB degradation, the peptide region containing the NLS is exposed in NF-κB and the nuclear import mechanism kicks in. NF-κB is translocated into the nucleus through the nuclear pores by association with the importin complex. \(^{42}\) Studies on the influence of cargo size for molecules that are transported into the nucleus through this mechanism have shown that the rate of transportation has been found to be inversely proportional to the molecular size of the cargo. Small cargo sizes result in easier transportation, with increased rates as compared to larger cargo sizes. Analysis of the size of the modeled oxidized structure shows that it is significantly more compact than the reduced form which would enable its easier entry through the nuclear pores. The size of the smallest box that can fit the entire molecule or the molecular volume was calculated using Deepview\(^ {43}\) for both the structures and it was found that the smallest box capable of containing the oxidized structure was 5032 Å\(^3\) smaller than that for the reduced form. This appreciable reduction in volume is expected to improve its ability to pass through the nuclear pore. We hypothesize that oxidation in the cytosol facilitates nuclear translocation of NF-κB, due to the smaller size of the oxidized form. However, it is difficult from this study to
determine unequivocally the extent to which the size reduction helps in the nuclear import mechanism.

Figure 3.5: Space-filling model of the oxidized and reduced forms of NF-κB

The observed decrease in volume from the surface and core of the protein shows that the volume of the DNA-binding cleft is also reduced in the oxidized form, making it constrained for DNA to enter and bind to NF-κB as a result of steric hindrance. It is exactly what is expected from the convergence of two subunits resulting in a decrease in the volume of the binding pocket, making it difficult for DNA to interact with the residues in the pocket and in fact making it impossible for DNA to enter the binding cleft. This demonstrates the need for Trx to reduce the disulfide bond in order to enable DNA-binding and transcription activation. The oxidized structure thus highlights the intricate and multi-faceted relationship between redox status and NF-κB activation, stressing that NF-κB responds to different stimuli in the cytoplasm and in the cell nucleus.
Figure 3.6: Comparison of the DNA-binding cleft of the reduced and oxidized forms (a)
Reduced form of NF-κB (PDB code 1NFK) with the bound DNA. The Cys 62 (Cys 59 in PDB
file) residues in the two subunits are shown in yellow. (b) Modeled structure of the oxidized
form of NF-κB showing the inter-subunit disulfide bond (in yellow).

3.5 Conclusion

The model of the oxidized form of NF-κB obtained from this study indicates that the flexible
linker region that connects the dimerization and DNA-binding domains provides flexibility to the
protein structure necessary for the relatively large movements upon oxidation and reduction. It is
known that disulfide bridges often play functional roles in proteins and studies have shown that
the enzyme thioredoxin acts as a regulatory switch of target proteins by reducing their disulfide
bonds. The structural data from this study suggests that the disulfide bond formation that results
in the closed conformation of the dimer is the form that recognizes and interacts with thioredoxin
in order to be reduced before it can bind to the DNA. After the Cys-62 residues that are involved in disulfide bonding are reduced, the DNA-binding regions of the two subunits which were in contact with each other move farther apart which enables its binding to DNA. The DNA then enters the cleft that is formed when NF-κB is reduced, where it can bind and start the transcription process. This proposed oxidized form of NF-κB from our computational studies thus provides structural explanation to the questions that were not answered by the previous experimental studies that showed NF-κB bound to either DNA or I-κB molecules. The snapshot of the oxidized model provides insight into an intermediate structure of the NF-κB activation pathway and also shows the importance of the redox signaling in controlling transcriptional activation. However, in a specific cell type not only are there different NF-κB dimers, but the large number of combinations between I-κB, DNA and NF-B dimers illustrates the complexity of the system and the limitations involved in modeling studies. Further experimental work such as NMR studies will be required to solve the intermediate structures and better understand the dynamic nature of the pathway.

3.6 References


36. The program SYBYL 7.0 is available from Tripos Inc., South Hanley Road, St. Louis, MO 63144, USA.


42. R. Fagerlund, L. Kinnunen, NF-kB Is Transported into the Nucleus by Importin α3 and Importin α4, J. Biol. Chem. 280 (2005) 15942-15951.
CHAPTER 4

MOLECULAR DOCKING AND ANALYSIS OF INTERACTIONS BETWEEN
VASCULAR ENDOTHELIAL GROWTH FACTOR (VEGF) AND SPARC PROTEIN

4.1 Abstract

The extracellular module of SPARC binds to vascular endothelial growth factor (VEGF) and inhibits VEGF-stimulated proliferation of endothelial cells. In an attempt to identify the binding site for SPARC on VEGF, we hypothesized that this binding site could overlap at least partially the binding site of VEGF receptor 1 (VEGFR-1), as SPARC acts by preventing VEGF-induced phosphorylation of VEGFR-1. To this end, a docking simulation was carried out using a predictive docking tool in conjunction with biological information, to obtain modeled structures of the VEGF-SPARC complex. The predicted structure of VEGF-SPARC complex indicates that the extracellular domain of SPARC interacts with the VEGFR-1 binding site of VEGF. Following molecular dynamics, side-chain interactions were identified at the protein interface that are predicted to contribute a large fraction of binding free energy. The identified interactions will be used for directing further mutagenesis studies to investigate their effect on the binding activity. The docking model obtained will provide a more detailed understanding of the protein-protein interactions and can be used as a basis for guiding future studies aimed at identifying VEGF receptor antagonists.
4.2 Introduction

Angiogenesis, the process of new capillary blood vessel growth from pre-existing vasculature, is a necessary physiological process in growth and development as well as in wound healing. Under normal circumstances angiogenesis occurs as a highly ordered series of events, spreading the vascular network only to the extent required by the demands of growing tissues\(^1\). However, in spite of the tight regulation, angiogenesis can occur not only in normal development and physiological processes but also in pathological processes such as tumor growth and metastasis.\(^2\)

The role of angiogenesis in tumor progression has been well established as a critical step in the transition of tumors from a dormant state to a malignant state through the supply of oxygen and nutrients to the proliferating cells.\(^3,4\) Many positively and negatively acting factors influence angiogenesis, which is a complex multi-component process involving the coordinated action of many growth factors and their receptors, cytokines, proteases, extracellular matrix proteins and adhesion molecules.\(^5,6\) Tumor cells induce new blood vessel growth by up-regulating growth factors such as VEGF, thereby overwhelming the effect of natural angiogenesis inhibitors.\(^7-9\)

Abnormal blood vessel growth in the eye is a leading cause for catastrophic loss of vision and extensive research has been done to understand the basic, underlying mechanisms of ocular angiogenesis.\(^10,11\) Vascular endothelial growth factor (VEGF)-related family of angiogenic factors are believed to play a central role in ocular angiogenesis as they regulate a wide variety of endothelial cell functions including cell survival, migration, differentiation and vascular permeability. Abnormal VEGF level causes proliferation of vessels into the retina leading to loss of vision.\(^12-14\) Conventional treatment methods have met with limited success and inhibition of growth factor-induced new vessel formation and targeting of pathological vessels seem to be the best choice for treating angiogenesis-related ocular diseases.\(^15,16\)
VEGF is the most well-characterized angiogenic and vascular permeability enhancing factor and it is upregulated in many different tumor types. It is a critical regulator of both physiological and tumor angiogenesis,\textsuperscript{17,18} and exerts its function by interacting with two high affinity tyrosine kinase receptors; \textit{fms}-like tyrosine kinase (Flt-1/VEGFR-1) and the kinase insert domain-containing receptor (KDR/VEGFR-2). Although these receptors are homologous and share architectural similarity they are significantly different in their functional properties.\textsuperscript{19-22} As a result, the action of VEGF and the function it mediates depends on the site of its expression, the type of receptors present and the signals they initiate.\textsuperscript{23} The Flt-1 gene, for instance, encodes for both the full-length receptor Flt-1 (VEGFR-1) and a soluble form designated sFlt-1. sFlt-1 is a known endogenous inhibitor of VEGF and a potent anti-angiogenic factor. sFlt-1 carries the VEGF-binding domain of Flt-1 as well as a unique intron-encoded 31-amino acid residue carboxyl-terminal sequence, but is devoid of the membrane proximal immunoglobulin-like domain, the membrane-spanning polypeptide, and the entire intracellular tyrosine kinase-containing region. It tightly binds VEGF with the same affinity as the full length receptor, suppressing its angiogenic activity.\textsuperscript{24,25} Normally a physiological angiogenic balance is maintained with the endogenous inhibitors of angiogenesis countering the effect of proangiogenic molecules and studies have identified many such molecules and their roles in inhibiting angiogenesis. The interactions and delicate balance between these different angiogenic and angiostatic factors are critical for regulating angiogenesis. Therefore, antivascular treatment approaches are targeted to interfere with the balance of these angiogenic and angiostatic factors.\textsuperscript{26,27}

The glycoprotein SPARC (Secreted Protein Acidic and Rich in Cysteine) has been found to act as an angiogenesis inhibitor by regulating the activities of growth factors like VEGF and
platelet-derived growth factor (PDGF).\textsuperscript{28-31} SPARC is a Ca\textsuperscript{+2}-binding, multifunctional glycoprotein belonging to the class of matrix-associated factors that mediate cell-matrix interactions. The primary structure of SPARC contains four unique domains with distinct functional activities and binds Ca\textsuperscript{+2} at both amino- and carboxyl-terminal domains. It has been shown to be involved in the regulation of important physiological processes through the modulation of cell-matrix and cell-growth factor interactions.\textsuperscript{32,33} Specific peptide regions have been identified in the protein that inhibit cellular proliferation and regulate the migration of endothelial cells. Recent studies have shown that SPARC binds to VEGF and inhibits VEGF-stimulated proliferation of endothelial cells. It is our hypothesis that this binding site could overlap at least partially the binding site of VEGFR-1, as SPARC acts by preventing VEGF-induced phosphorylation of VEGFR-1.\textsuperscript{34} In order to understand the basis for the multifunctional role of SPARC,\textsuperscript{35-37} further knowledge regarding its interaction with growth factors, especially VEGF, will be useful.

Despite the available individual structural information regarding VEGF and SPARC, the atomic details of their complex structure and the exact structural domains that mediate their interactions and subsequent biological effects still remain unclear. Computational methods can be useful in elucidating the interactions between these molecules and in our present study, molecular modeling and protein-protein docking simulations are carried out towards the twin goals of, predicting their complex structure to understand their molecular impact on angiogenesis, and identifying specific residues involved in the interaction towards designing antagonists that can inhibit the effect of VEGF to prevent the abnormal growth of new blood vessels.
The growing number of individual structures in the crystallographic databases and the relatively small number of solved complexes has made predictive docking an important theoretical method. Protein-protein complexes are especially important since these reactions are biologically abundant in nature and it would be desirable to understand them in detail. It has been estimated that 70% of proteins function through multiprotein complexes in yeast, although the number of interactions in humans is difficult to estimate. Because of the inherent difficulty in obtaining crystal structures of protein complexes, docking techniques could be used to predict the complex structures. Docking strategies usually rely on a two-stage approach: first, to generate a set of possible orientations of the two docked proteins and then score them in a way that the native complex will be ranked highly. In this study, the predictive docking tool FT Dock was used to perform the docking simulations through rigid-body docking of two biomolecules in order to predict their correct binding geometry. FT Dock and RPScore are part of the 3D-Dock suite which was obtained from the Biomolecular Modelling Laboratory. The docking procedure involves performing a global rigid-body search of rotational and translational space, and scoring each potential structure on shape and electrostatic complementarity. The RPScore then ranks each possible complex based on the observed residue-residue contacts, according to statistical potentials derived from the likelihood of contacts between residues of different types from solved protein-protein complexes taken from SCOP. The complexes ranked by this method are further screened using biochemical information and evaluated to select the best model. Docking is used in concert with experimental techniques, including site-directed mutagenesis to identify the correct complex structure. The extensive rigid-body docking and the use of structural and experimental data to filter the results, is expected to produce a reasonable model of the complex. The final complex structure is then studied to analyze the inter-molecular
contacts in terms of hydrogen bonding, salt-bridges, electrostatic and hydrophobic interactions. This general procedure is used in the present study to identify specific residue level interactions between the proteins and the results will be used to direct further mutagenesis studies to investigate their effects on binding activity. The goal of our research is to create a detailed model of the complex and offer a structural interpretation for the available experimental data.

4.3 Methods

The procedure to obtain the structure of docked complexes involved the following steps and is shown in Fig. 4.1: (i) A global scan of translational and rotational space of the two proteins using FTDock, based on the surface complementarity and an electrostatic filter; (ii) An empirical scoring of the obtained complexes using RPScore, based on residue level pair potentials; (iii) Evaluation of the scored complexes to filter the best conformation based on available experimental data; and (iv) Energy minimization and molecular dynamics to refine the structure and to improve the steric and electrostatic interactions between side-chains at the interface.

Figure 4.1: Protein-Protein Docking Procedure
4.3.1 Protein-protein docking

There are several programs available for protein-protein docking that attempt to predict the structure of docked complexes when the coordinates of the components are known. Most of the algorithms are based on rigid-body docking methods, in which the larger protein is kept fixed and the smaller protein is rotated and translated to find the best geometric fit. In this study, FTDock 2.0 was selected for performing the docking simulations as it uses fourier transform to rapidly evaluate the shape complementarities and also it has various post-docking processing methods to score the resultant complexes, including scoring based on electrostatics and experimental data. The program has good predicting ability, as the root mean standard distance of the obtained complexes can be within 1.0 Å from known crystal structures for cases where biological information is available. The primary advantage over other programs is that FTDock allows filtering of thousands of docked complexes to a manageable extent based on available biochemical information.

The starting unbound structures for the docking were obtained from the PDB database, the structure of VEGF bound to the second domain of Flt-1 receptor (PDB code: 1FLT) and the structure of SPARC (PDB code: 1BMO), both with resolutions of 3.0 Å or better obtained by X-ray diffraction. Starting from the known crystal structures, the end groups and disordered residues on both the proteins were first fixed using the Biopolymer module in Sybyl 7.0. The subsequent structure of SPARC was docked on to VEGF after removing the Flt-1 receptors from the starting structure. Before docking, the coordinates of proteins were preprocessed to remove hydrogen atoms and alternative atom records. Each molecule was digitized onto a 274 x 274 x 274 grid with a 0.7 Å grid unit and the surface thickness was set to 1.3 Å. During the docking simulation the larger of the two molecules, SPARC was held static and VEGF was translated and
rotated with respect to SPARC to explore all possible orientations. The rotational degrees of freedom were explored by rotating the VEGF molecule in 12° angle step size and three rotations at each step were retained. The shape complementarity of the two molecules was evaluated from the overlap of the two grid functions. The docking run, which results in 10000 docked complexes, was performed with the inclusion of electrostatic scoring for excluding false positive complexes. All the docking simulations presented in this study were performed on IBM-p655 cluster comprised of 32 processors. However, the FTDock program available from the developers is not parallelized for multiple processors and therefore the capability of the resources used did not reflect on the computational time taken for the simulation.

4.3.2 Rescoring of complexes

The subsequent step to docking was to reduce the number of complexes in the list that need to be considered in order to select the complex that will be closest to the true structure. The complexes obtained from the initial docking run were ranked based on the surface complementarity score for each of the complex structure. The complexes were rescored using RPScore, which was available as part of the 3D-Dock suite. The output from the docking run was sorted and ranked based on this alternative ranking, using the residue-residue pair potential scores for the complex structures. This scoring scheme based on evidence from actual protein interfaces is defined as the log fraction of the actual frequency and the expected frequency of occurrence, and the value of the score for each pair is a measure of the likelihood that the particular pair occurs and thus helps to quantify the probability of a complex’s existence.41

4.3.3 Filtering using experimental data

Biochemical data was used to further filter the docking results, using the residues on the surface of the interacting proteins as candidates. The results obtained from RPScore were filtered based
on the best documented experimental information that SPARC and the peptide 4.2 from the extra cellular region of SPARC bind to VEGF and inhibit VEGF induced proliferation of endothelial cells. Receptor activation studies have also shown that SPARC could prevent VEGF induced phosphorylation of VEGFR-1, by selectively blocking the activity of this receptor. These data suggest that the peptide 4.2 (aa 254-274) corresponding to the C-terminal region of SPARC should be in proximity to the receptor binding face of VEGF, which is primarily composed of acidic residues (aa 63-67) and some basic residues (aa 82-86) which mediate interaction with VEGFR-1. This biological information was used as the main constraint for filtering and selecting the docked complex. Additionally, it is known that residues from both the VEGF dimers are involved in receptor binding interactions with VEGFR-1 and in order to bind competitively to VEGF and inhibit the phosphorylation of VEGFR-1, the interacting surface between the two proteins should be sufficiently large and involve all the key residues. This was not considered as a necessary criterion, but used only to rank complexes that matched both the criteria. The top ranked 100 complexes obtained from this filter were visually analyzed and the residues implicated by experimental results as having an influence on the interaction were used to guide the manual selection of final docked complex. A model structure for the docked complex was selected that was most compatible with the available experimental information.

4.3.4 Molecular dynamics simulations

The best structural model for the ternary complex of VEGF-SPARC obtained from the docking procedure was subjected to MD simulation to refine the protein interface. However, no explicit constraint functions were used to maintain the initial docking contacts during the simulation. The structures were first energy minimized using 1000 steps of steepest descent and 2000 steps of conjugate gradient minimization using Kollman all-atom force field implemented in SYBYL.
A distance dependent dielectric function was used with the dielectric constant set to 1 and the non-bonded cutoff was set to 8 Å. Energy minimization with classical force field can be used to remove unrealistically close steric clashes and large deviations from ideal geometry resulting from the conformational changes of amino acid side chains after docking, but molecular dynamics simulation is required to improve rotamer distributions. This energy minimized structure was used as the starting structure for the MD simulation. All MD simulations were performed with the AMBER 7.0 molecular simulation package. A 3 ns (nanosecond) simulation was performed under constant pressure conditions using the parm99 force field to describe the interactions between the protein atoms. As in the minimization, the dielectric constant was set to 1.0 was used and an 8 Å residue charge group based cutoff for nonbonded interactions was used, with the nonbonded pairlist updated every 15 steps. This cutoff was chosen to reduce the computational time involved due to the large size of the system. The bonds involving hydrogen atoms were constrained using SHAKE algorithm. The simulation was using a 2 fs time step for a total simulation time of 3 ns. The final conformation obtained at the end of the MD simulation was used for identifying specific interactions at the interface, computing amino acid residue distances and other calculations.

4.4 Results and Discussion

The complex structure of VEGF-SPARC which was derived from protein-protein docking simulation must be consistent with the available experimental information and also should reflect the nature of protein-protein interactions in general. Predicting protein-protein interactions is inherently challenging owing to the difficulty in modeling the many forces that contribute to these interactions. This leaves the burden of excluding false positives from the docking results and ascertaining whether the model obtained is reliable by using accurate scoring and filtering
techniques. Apart from using surface complementarity and electrostatic filter, residue pair potentials and biochemical data were also included to score the docking orientations, as it has been shown to produce more accurate results than using geometric fit and electrostatic energy alone.\textsuperscript{41,43} The most favorable solution obtained by this method was then refined through molecular dynamics to get the final docked model, shown in figure 4.2, which was used to analyze the interactions at the protein interface.

\textbf{Figure 4.2:} Cartoon representation of the complex structure of VEGF-SPARC obtained through docking simulation: VEGF monomers are colored green and blue, while SPARC is colored red.

\textbf{4.4.1 Scoring and filtering the docked conformations}

In order to upgrade these models to reliable predictions, which could be used with confidence for further experimental and computational work, refinement using biological data is done. Any docking program attempting to find a complex structure for two given molecules based on surface complementarity and geometric fitting would invariably return several docking poses between the two molecules.\textsuperscript{54} However, a large percentage of the docking orientations would be
biologically irrelevant and can be easily eliminated through simple visual inspection, if some experimental information regarding their interaction is known. In this study, the presence of specific experimental information regarding the binding of SPARC to VEGF proved very useful as a means of filtering the docking results. The list of possible orientations was narrowed down to include only biologically relevant structures by incorporating experimental information that the binding of SPARC to VEGF is mediated specifically through the peptide 4.2 region of SPARC.\textsuperscript{34,55,56} Additionally, we also took into account the hypothesis that the binding site of SPARC on VEGF should overlap, at least partially, the binding site of VEGFR-1. A distant constraint was therefore applied to filter the structures to include only the orientations where the peptide 4.2, comprising of amino acid residues 255-274, was within 4 Å distance from the VEGFR-1 binding site. Only the complex structures that correlated well with this experimental information were retained and the top 100 orientations were analyzed and compared with one another to select the final docked conformation. Apart from this biochemical data, the selection of the final model for the docked complex was also based on factors like the area of surface contact, extent of interactions present and stability of the model. This led to the model structure of the VEGF-SPARC complex (Figure 4.2) in which there are several identified intermolecular hydrogen bonds between the two molecules, apart from hydrophobic and other long range electrostatic interactions.

4.4.2 Identification of side chain interactions

Long-range non-specific electrostatic interactions as well as short-range electrostatics like hydrogen bonds and salt bridges are crucial for complex stability.\textsuperscript{57} To determine the mode of action, information about which amino acid residues come into contact when the complex is formed is important to supplement the data obtained experimental methods. Also as mentioned
above, the information on amino acid residues is also important to validate our working hypothesis of an overlapping binding site for SPARC and VEGFR-1, in order for SPARC to inhibit the VEGF-stimulated phosphorylation of VEGFR-1. Results from previous mutagenesis studies have shown specific regions in VEGF responsible for receptor binding interactions with VEGFR-1 and VEGFR-2. The region composed of predominantly acidic residues (aa 63-67) mediates binding to VEGFR-1 and another region composed of basic residues (aa 82-86) mediates binding to VEGFR-2. These oppositely charged surface regions are present at distal ends of the monomer, but since VEGF molecules exist as disulfide-linked homodimers these two regions are in close proximity in the dimeric form creating a cluster of receptor binding determinants at each end of the VEGF molecule. Apart from this cluster of residues, the VEGFR-2 binding determinants form two other hot spots: Phe 17' and Gln 79; and Ile 46, Glu 64' and Ile 83', which are present on the same face that is responsible for receptor-binding in VEGF.

Figure 4.3: VEGF receptor binding interactions; red solid representation corresponds to the acidic residues that mediate VEGFR-1 binding (aa 63-67) and blue solid representation corresponds to the basic residues that mediate VEGFR-2 binding (aa 82-86).
In our efforts to identify the key residues that drive the interaction between VEGF and SPARC and stabilize their complex, the side chain interactions between the following amino acids, listed in Tables 4.1 and 4.2, were identified as crucial for binding activity and these binding hot spots will be used to guide site-directed mutagenesis studies.

**Table 4.1: Residues at the interface responsible for hydrogen-bonding interactions** (Range 2.0-3.2 Å)

<table>
<thead>
<tr>
<th>Chain</th>
<th>Position</th>
<th>Residue</th>
<th>Atom</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>259</td>
<td>ASP</td>
<td>OD1</td>
<td>2.66</td>
</tr>
<tr>
<td>A</td>
<td>259</td>
<td>ASP</td>
<td>OD2</td>
<td>2.68</td>
</tr>
<tr>
<td>A</td>
<td>276</td>
<td>LYS</td>
<td>NZ</td>
<td>2.86</td>
</tr>
<tr>
<td>A</td>
<td>276</td>
<td>LYS</td>
<td>NZ</td>
<td>2.76</td>
</tr>
<tr>
<td>A</td>
<td>151</td>
<td>ARG</td>
<td>NH1</td>
<td>3.20</td>
</tr>
<tr>
<td>A</td>
<td>201</td>
<td>ARG</td>
<td>NE</td>
<td>3.17</td>
</tr>
<tr>
<td>A</td>
<td>204</td>
<td>GLU</td>
<td>OE2</td>
<td>3.05</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>O</td>
<td>2.97</td>
</tr>
<tr>
<td>A</td>
<td>272</td>
<td>CYS</td>
<td>O</td>
<td>3.05</td>
</tr>
</tbody>
</table>

Chain A denotes the SPARC molecule; chains V and W denote the VEGF dimers.

**Table 4.2: Residues at the interface responsible for hydrophobic interactions** (Range 2.0-3.8 Å)

<table>
<thead>
<tr>
<th>Chain</th>
<th>Position</th>
<th>Residue</th>
<th>Atom</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CB</td>
<td>3.63</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CD2</td>
<td>3.8</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CA</td>
<td>3.58</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>C</td>
<td>3.65</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CD2</td>
<td>2.6</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CD2</td>
<td>3.4</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CD2</td>
<td>3.58</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CD2</td>
<td>2.64</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CG</td>
<td>3.77</td>
</tr>
<tr>
<td>A</td>
<td>271</td>
<td>GLY</td>
<td>C</td>
<td>3.77</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CD1</td>
<td>3.09</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CD1</td>
<td>2.31</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CD1</td>
<td>2.86</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CD1</td>
<td>3.62</td>
</tr>
<tr>
<td>A</td>
<td>258</td>
<td>LEU</td>
<td>CG</td>
<td>3.51</td>
</tr>
<tr>
<td>A</td>
<td>277</td>
<td>GLN</td>
<td>CD</td>
<td>3.78</td>
</tr>
</tbody>
</table>

Chain A denotes the SPARC molecule; chains V and W denote the VEGF dimers.
The side-chain interactions involving these residues at the interface of VEGF and SPARC contribute a large fraction of binding free energy, highlighting their importance in stabilizing the protein complex. These residues define finite areas on the surface of each protein, corresponding to the interaction surfaces.
c)

**Figure 4.4:** Amino acid residues involved in side chain interactions at the VEGF-SPARC binding interface. Residues a) Asp 63; b) Gln 79; and c) His 86 from VEGF stabilize the complex formation.

The hydrogen bonding interactions involving several key residues in the binding hot spots are consistent with our hypothesis that SPARC binds to VEGF on the same receptor binding face that mediates its interaction with VEGFR-1. The prevalence of several non-polar amino acids is a characteristic trait of docking interfaces.\(^{52}\) Although docking methods are not sufficiently accurate to predict whether or not two proteins actually interact with each other, they can sometimes correctly identify the interacting surfaces between two structurally defined subunits when right methods are used.

**4.4.3 Computational alanine scanning**

The role of individual amino acid side chains in stabilizing the complexes was further probed by computational alanine scanning studies, which identifies residues that are important for the
stabilization of the complex, by determining the change in the free energy of binding when various residues in the wild type protein was mutated to alanine.\textsuperscript{59,60} The results from the alanine scanning experiments, listed in Tables 4.3 and 4.4, correlated well with the docking simulation results.

**Table 4.3:** Results of the predicted contributions of residues through virtual alanine scanning

<table>
<thead>
<tr>
<th>PDB #</th>
<th>Chain</th>
<th>Int_ID</th>
<th>Δ∆G(bind)</th>
<th>ΔG(partner)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>W</td>
<td>1</td>
<td>1.46</td>
<td>-0.21</td>
</tr>
<tr>
<td>17</td>
<td>W</td>
<td>1</td>
<td>1.70</td>
<td>0.06</td>
</tr>
<tr>
<td>21</td>
<td>W</td>
<td>1</td>
<td>1.97</td>
<td>-0.39</td>
</tr>
<tr>
<td>63</td>
<td>W</td>
<td>1</td>
<td>1.24</td>
<td>-0.48</td>
</tr>
<tr>
<td>258</td>
<td>A</td>
<td>1</td>
<td>1.70</td>
<td>0.74</td>
</tr>
<tr>
<td>259</td>
<td>A</td>
<td>1</td>
<td>1.19</td>
<td>-0.77</td>
</tr>
<tr>
<td>267</td>
<td>A</td>
<td>1</td>
<td>1.17</td>
<td>1.42</td>
</tr>
<tr>
<td>276</td>
<td>A</td>
<td>1</td>
<td>1.15</td>
<td>-0.92</td>
</tr>
</tbody>
</table>

Chain A denotes the SPARC molecule; chain W denotes the VEGF monomer.

**Table 4.4:** Results of the predicted contributions of residues through virtual alanine scanning

<table>
<thead>
<tr>
<th>PDB #</th>
<th>Chain</th>
<th>Int_ID</th>
<th>Δ∆G(bind)</th>
<th>ΔG(partner)</th>
</tr>
</thead>
<tbody>
<tr>
<td>148</td>
<td>A</td>
<td>1</td>
<td>1.25</td>
<td>2.16</td>
</tr>
<tr>
<td>151</td>
<td>A</td>
<td>1</td>
<td>2.48</td>
<td>2.52</td>
</tr>
<tr>
<td>255</td>
<td>A</td>
<td>1</td>
<td>1.63</td>
<td>0.20</td>
</tr>
<tr>
<td>258</td>
<td>A</td>
<td>1</td>
<td>1.66</td>
<td>0.74</td>
</tr>
<tr>
<td>272</td>
<td>A</td>
<td>1</td>
<td>1.01</td>
<td>5.33</td>
</tr>
<tr>
<td>79</td>
<td>V</td>
<td>1</td>
<td>1.48</td>
<td>0.16</td>
</tr>
<tr>
<td>86</td>
<td>V</td>
<td>1</td>
<td>1.08</td>
<td>-0.61</td>
</tr>
<tr>
<td>89</td>
<td>V</td>
<td>1</td>
<td>4.03</td>
<td>-0.43</td>
</tr>
<tr>
<td>90</td>
<td>V</td>
<td>1</td>
<td>1.67</td>
<td>-0.52</td>
</tr>
<tr>
<td>91</td>
<td>V</td>
<td>1</td>
<td>0.96</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Chain A denotes the SPARC molecule; chain V denotes the VEGF monomer.

(Column 1) PDB#, number of mutated residue in the pdb file; (Column 2) Chain, pdb chain identifier; (Column 3) Int_ID, measure of whether a residue is interacting directly (1) or not interacting directly but is buried upon binding (0); (Column 4) Δ∆G(bind), predicted change in binding free energy upon alanine mutation; (Column 5) ΔG(partner), predicted change in protein stability of the mutated partner upon alanine mutation.
The experimental studies carried out previously showed that the peptide 4.2 region of SPARC inhibited VEGF-induced cell proliferation and the model of the VEGF-SPARC complex obtained from our docking studies showed that the residues in this region play a significant role in binding stability. The results from computational alanine scanning confirm that these residues are important for the stability of the complex. Positive values of $\Delta\Delta G$ means that the alanine mutation is predicted to destabilize the complex and negative values indicate a stabilizing effect. Computational mutation of Asp 63, Gln 79, His 86 and Gln 89 in VEGF as well as several residues in the region from 255-274 in SPARC, to alanine had unfavorable effects on the stability of the complex. This study indicates that the enthalpic contribution from the desolvation of amino acids, formation of novel H-bonds, van der Waals and electrostatic interactions involving these residues contribute to a favorable free energy of interaction between VEGF and SPARC, and offset the decrease in entropy from the loss of translational and rotational degrees of freedom upon binding.

It is also interesting to note that several important hydrogen bonding and hydrophobic interactions observed in our proposed model for VEGF-SPARC complex, not only validate our hypothesis of an overlapping binding with VEGFR-1, but also indicate the possibility of an overlapping binding site with VEGFR-2, as several of the critical residues responsible for VEGFR-2 binding are involved in SPARC recognition as well. The hydrogen bonding interactions involving residues Gln 79 and His 86, and hydrophobic interactions involving residues Phe 17, Arg 82 and Ile 83 which are also crucial for VEGFR-2 binding strongly support the involvement of the same hot spot of residues on the VEGF surface, in mediating its interaction with the receptors as well as with molecules such as SPARC. The current model agrees well with the experimental data that the region of residues 255-274 in SPARC possess
anti-angiogenic functions through its binding with VEGF. While the present docking model supports a non-selective inhibition of the activity of VEGF receptors, it does not explain the results obtained from earlier studies which found that SPARC did not inhibit VEGF-mediated activation of VEGFR-2.\textsuperscript{34} Therefore, further studies using site-directed mutagenesis and other experimental methods will be able to verify the binding sites identified by our docking model.

4.5 Conclusion

In the current study, we have demonstrated the application of protein-protein docking simulation to build a complex structure of VEGF-SPARC starting from unbound proteins using the program FTDock. Starting from unbound structures computer docking simulation can be used to build a set of atomic models of complexes, one of which will be close to the native complex structure. To drive the docking simulation towards a limited set of solutions with biological significance, a filter, incorporating general considerations of protein-protein interactions, specificity of the selected protein and results from experimental work was employed. By applying proper filtering and scoring methods, it is possible to select the right structure from the docking results. We have used electrostatics, residue pair potentials and biochemical information in this study to filter and rank the docked models and build a reliable model of the complex structure. After filtering, the final model of the complex was selected that agreed best with the biological data and it was refined using molecular dynamics, to analyze the interactions and identify hot spot residues. The hot spots in the protein-protein interface, which are small regions that are crucial to binding, can be targeted by small molecules to mimic the protein-protein interactions. Our study has enabled us to put forward a model in which SPARC binds to VEGF near its VEGFR-1 binding domain. This model can be used for future experimental and computational studies to draw biological and functional conclusions. Using mutagenesis studies to be carried out based on the interactions
identified in this study it is possible to prove the exact mechanism of SPARC binding to VEGF. We have thus constructed an experimental based model of the complex using computational docking in conjunction with biological information to understand the molecular basis of their interaction. Knowledge of the interactions and the important binding residues will help to develop new strategies to target VEGF and inhibit angiogenesis. This proposed model is primarily aimed at predicting the orientation of SPARC against VEGF, identifying important interactions based on that prediction and prioritizing experimental studies like site-directed mutagenesis. However, to improve the structural resolution and elucidate details at the atomic level, more experimental work is needed.

4.6 Acknowledgements

This research was supported by Knights Templar grant through Drs. B. Ambati and J. Ambati. Computing resources were supported by Research Computing Center (RCC) at the University of Georgia.

4.7 References


45. The program SYBYL 7.0 is available from Tripos Inc., South Hanley Road, St. Louis, MO 63144, USA.


56. J.C. Yost, E.H. Sage, Specific interaction of SPARC with endothelial cells is mediated through a carboxyl-terminal sequence containing a calcium-binding EF hand, J. Biol. Chem. 268 (1993) 25790-25796.
57. L. Pauling, M. Delbruck, The nature of the intermolecular forces operative in biological processes, Science 92 (1940) 77-79.
58. Y.A. Muller, H.W. Christinger, The crystal structure of vascular endothelial growth factor (VEGF) refined to 1.93 A resolution: multiple copy flexibility and receptor binding, Structure 5 (1997) 1325-1338.
CHAPTER 5
CONCLUSIONS AND DIRECTIONS

Successful application of ligand and structure-based molecular modeling methods have been made to various problems in several systems, to build structural models and make functional predictions. Computational methods have been used in a wide variety of biomedical and drug discovery projects till date to understand and predict the properties of small molecules as well as investigate the behavior of complex macromolecular systems. In this research work, the structural properties of three specific systems were examined through computer simulation methods to enhance existing knowledge about these systems.

QSAR Studies

**Specific aim:** Develop predictive QSAR models for MNVP⁺ analogs as choline acetyltransferase (ChAT) inhibitors, using which the activities of newly designed analogs can be predicted, prior to synthesis. 3D-QSAR models were derived, employing the CoMFA and CoMSIA methods consisting of a set of 21 compounds with known activity data. The model developed gives insight into the different field contributions around the molecules and their influence on the overall activity. Three-dimensional descriptors and log P values for the compounds were used to allow correlation of three-dimensional structures and biotransport of the molecules with their biological activity. The QSAR models were assessed for their predictive abilities using a test set of four compounds which were not used in the development of the model. The CoMSIA model performed better than the CoMFA models, possibly because of the inclusion of additional
descriptors that aid pharmacophore element searching and correlate better with activity. Mapping of molecular surface properties of the compounds can lead to a better understanding of the receptor surface and binding mode for these ChAT inhibitors. The proposed pharmacophore model from our studies can be used to further refine the structural features and aid in increasing the activity and biotransport of the molecules. This study provides further insight to support ligand-based design of bioactive compounds and demonstrate the power of combining 3D-QSAR methods with experimental work to explore the binding mechanism and structural properties of for ligands where these are unknown.

NF-κB: Oxidized form

Specific aim: Model the oxidized form of NF-κB to understand its role in activation pathway and cellular entry of the transcription factor. The primary intention of the project was to build the oxidized model of NF-κB, as the structures available so far have paid attention only to the reduced form of NF-κB bound to DNA or I-κB molecules and there is no snapshot of the oxidized structure available. A preliminary analysis of the NF-κB structure indicated that the flexible ten amino acid linker region that connects the dimerization and DNA-binding domains is responsible for producing the domain movements. A model for the oxidized form was built through systematic variation of torsion angles in the linker region keeping the secondary structural features and energetic requirements in mind. The oxidized structure was formed through a disulfide bridge involving Cys-62 residues in the DNA-binding region of NF-κB. The structure was then refined through molecular dynamics to obtain the final model for the oxidized form of NF-κB. The domain movement observed in the structure agreed well with protein domain movements in general and the structure was able explain various aspects of NF-κB signaling pathway including nuclear entry and subsequent reduction by Trx before it binds to
DNA. The findings from this study help to determine how the oxidized form of NF-κB is more compact than the reduced form, thus facilitating nuclear transport of the transcription factor. Also, the oxidized form of NF-κB is the structure that recognizes and interacts with Trx. The reduction by Trx is necessary for DNA-binding of NF-κB, as the disulfide bridge in the oxidized form prevents the DNA from entering the cleft and binding to NF-κB.

**VEGF-SPARC complex structure**

Specific aim: Predict the complex structure of vascular endothelial growth factor (VEGF) with SPARC through docking simulation and identify crucial binding interactions. The binding interaction between VEGF and SPARC was explored by building their complex structure through molecular docking studies. Protein-protein docking simulation was used to build the complex structure starting from their unbound structures, using shape complementarity and electrostatics. The docked models were scored using residue pairing potentials and experimental data was used to further filter the results to build an experimental-based model of the complex. The model that agreed best with biological information was selected and refined using molecular dynamics studies to analyze the interactions at the molecular interface. Residues that were predicted to contribute to a large fraction of binding free energy were identified and confirmed using computational alanine scanning studies. The identified binding residues can be used in future mutagenesis studies to confirm the docking predictions. Comparison with experimental results was favorable showing that the model is biologically relevant. The proposed model of the complex structure can be used to understand the molecular mechanism involved in the inhibition of VEGF-induced angiogenesis by SPARC. This model can also be used a basis for guiding studies aimed at designing small molecule inhibitors based on the identified hot spots in the protein interface.
All computational methods depend one way or another on experimental data, starting from molecular dynamics force fields that are derived from protein and peptide data or docking simulations that use experimental data to filter and validate the results. Unfortunately, there are lot of errors involved in experimental data and protein structures used in modeling. Therefore, verification of the data used in modeling procedures is a prerequisite for good results.

It is also very important to note that this study is only as powerful as the methods it builds upon. In particular, docking is still a highly active area of research with much work remaining to be done on model flexibility, solvent simulation, and force field optimization. Similarly, model building of the protein complex remains a difficult procedure requiring great care, making it difficult to make accurate predictions. Awareness of the strengths and limitations of different experimental and computational approaches is critical in the success of a drug discovery effort.