RATIONAL DESIGN OF SMALL-MOLECULE INHIBITORS OF PROTEIN-PROTEIN INTERACTIONS: APPLICATION TO THE ONCOGENIC C-MYC/MAX INTERACTION

by

Lidio Marx Carvalho Meireles

B.S., Computer Science, Universidade Federal de Sergipe, Brazil, 2003

M.S., Computer Science/Bioinformatics, Kyoto University, Japan, 2006

Submitted to the Graduate Faculty of

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Lidio Marx Carvalho Meireles

It was defended on

July 26th, 2011

and approved by

Dr. Ivet Bahar, Professor, Computational and Systems Biology

Dr. Gabriela Mustata, PharmacoInformatics Specialist, Computational and Systems Biology

Dr. Gordon Rule, Professor, Biological Sciences, Carnegie Mellon University

Dr. Edward Prochownik, Professor, Molecular Genetics and Developmental Biology

Dr. Chakra Chennubhotla, Assistant Professor, Computational and Systems Biology

Dissertation Advisor / Co-Advisor: Dr. Ivet Bahar / Dr. Gabriela Mustata

RATIONAL DESIGN OF SMALL-MOLECULE INHIBITORS OF PROTEIN-PROTEIN INTERACTIONS: APPLICATION TO THE ONCOGENIC C-MYC/MAX INTERACTION

Lidio Marx Carvalho Meireles, PhD

University of Pittsburgh, 2011

ABSTRACT

Protein-protein interactions (PPIs) constitute an emerging class of targets for pharmaceutical intervention pursued by both industry and academia. Despite their fundamental role in many biological processes and diseases such as cancer, PPIs are still largely underrepresented in today's drug discovery. This dissertation describes novel computational approaches developed to facilitate the discovery/design of small-molecule inhibitors of PPIs, using the oncogenic c-Myc/Max interaction as a case study.

First, we critically review current approaches and limitations to the discovery of small-molecule inhibitors of PPIs and we provide examples from the literature.

Second, we examine the role of protein flexibility in molecular recognition and binding, and we review recent advances in the application of Elastic Network Models (ENMs) to modeling the global conformational changes of proteins observed upon ligand binding. The agreement between predicted soft modes of motions and structural changes experimentally observed upon ligand binding supports the view that ligand binding is facilitated, if not enabled, by the intrinsic (pre-existing) motions thermally accessible to the protein in the unliganded form.

Third, we develop a new method for generating models of the bioactive conformations of molecules in the absence of protein structure, by identifying a set of conformations (from different molecules) that are most mutually similar in terms of both their shape and chemical features. We show how to solve the problem using an Integer Linear Programming formulation of the maximum-edge weight clique problem. In addition, we present the application of the method to known c-Myc/Max inhibitors.

Fourth, we propose an innovative methodology for molecular mimicry design. We show how the structure of the c-Myc/Max complex was exploited to designing compounds that mimic the binding interactions that Max makes with the leucine zipper domain of c-Myc.

In summary, the approaches described in this dissertation constitute important contributions to the fields of computational biology and computer-aided drug discovery, which combine biophysical insights and computational methods to expedite the discovery of novel inhibitors of PPIs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I gratefully appreciate the mentorship of my advisors Prof. Ivet Bahar and Dr. Gabriela Mustata. Thanks to their advising and support, I experienced a period of exceptional productivity in my academic life.

I am indebted to the dissertation committee members (Prof. Edward Prochownik, Prof. Gordon Rule and Prof. Chakra Chennubhotla) for their insightful questions and constructive comments and suggestions which helped me improve this work.

I also thank Prof. Edward Prochownik and his team for experimentally testing "my molecules" with high enthusiasm, and Prof. Peter Wipf for promptly providing access to the molecules of the Center for Chemical Methodologies and Library Development.

I am grateful to Dr. Christian Lemmen from BioSolveIT for supporting my research with the software ReCore and providing useful comments and suggestions.

Finally, I thank my dear wife Chie and my parents Claudio and Livia for their love and support.

I dedicate my PhD to my son, Alex.

1.0 INTRODUCTION

1.1 BACKGROUND AND SIGNIFICANCE

Biological processes involve the concerted interaction of biomolecules, among which protein-protein interactions (PPIs) play a central role. Knowing how biomolecular interactions enable biological processes helps researchers identify the underlying causes of diseases and ultimately develop targeted therapeutic strategies, like designing drugs to modulate the interaction between two proteins or the ligand-binding properties of a specific enzyme.

The ubiquitous role of PPIs in mediating biological processes in the cell and their involvement in a large number of diseases, such as cancer [1-3], make them attractive targets for pharmaceutical intervention. Traditionally, however, it was largely believed that finding small-molecule drugs to modulate PPIs would be nearly impossible, as protein-protein interfaces are generally large and devoid of well-defined cavities that can accommodate a small molecule – like those typically found on enzymes [4,5]. This view is perhaps reflected on the fact that most drugs approved by the Food and Drug Administration (FDA) target enzymes (kinases and proteases) and receptors (G-protein coupled receptors (GPCRs), ligand-gated ion channels and nuclear hormone receptors) [6]. Those 'traditional' drug targets (i.e., enzymes and receptors) usually present large cavities to which endogenous small molecules bind, and the drugs are

designed to compete for those same sites, often replacing the endogenous ligand by a synthetic analog [6].

Despite the skepticism, an increasing number of scientific publications reporting the discovery of new potent small-molecule modulators of PPIs [4,5,7-10] have demonstrated the feasibility of targeting PPIs with small molecules. The perception that PPIs are 'undruggable' has therefore changed; nowadays, PPIs are considered challenging but feasible drug targets offering great opportunities.

Some important lessons can be learned from the past two decades of research. Mutagenesis and structural studies have shown that, within the large contact surfaces of PPIs, discrete pockets of stronger attractions, so-called '*hotspots*', comprise sites that can be targeted by small molecules [11]. Moreover, the plasticity of the protein interface, which in many cases adapts to binding different partners, provides the opportunity, upon binding of a ligand, for formation of new sub-pockets that may not be expected from the examnination of X-ray structures alone [12,13]. As discussed in Chapter 2.0 new drug discovery paradigms emerged (e.g., Fragment-Based Drug Discovery) and existing methodologies have been adapted to these paradigms to possibly increase productivity in drug discovery, particularly for those difficult targets like PPIs.

Apart from the innumerous therapeutic applications, small-molecule modulators of PPIs also constitute valuable tools in Chemical Genetics studies, as probes to interrogate the function of PPIs participating in unknown biomolecular pathways or networks [14,15], which, in turn, facilitates the development of new targeted therapies. The ability to efficiently discover/design small-molecule modulators of PPIs is therefore of utmost importance to biomedical research.

This dissertation builds on and expands state-of-the-art computational methods, and proposes new protocols driven by biophysical insights, to accelerate the discovery of new inhibitors of PPIs. Two molecular modeling strategies are proposed: one based on known ligands, and the other based on the structure of the protein-protein complex. In addition, advances in modeling protein flexibility using Elastic Network Models are emphasized. As a study case, we apply the proposed strategies to design inhibitors of c-Myc/Max protein interaction, an important oncogenic target.

1.2 TARGETING C-MYC/MAX INTERACTION FOR CANCER THERAPY

c-Myc is a bHLH-Zip (basic Helix-Loop-Helix Leucine Zipper) transcription factor that regulates the expression of a large number of genes that collectively promote cell proliferation [16,17]. The active form of c-Myc exists as a heterodimer with another bHLH-Zip protein, Max [18-22]. The c-Myc/Max complex is necessary for transcriptional activation of target genes by binding the so-called Enhancer Box sites, “E-box” (CAC/GTG) on DNA [19,21,22] (Figure 1).

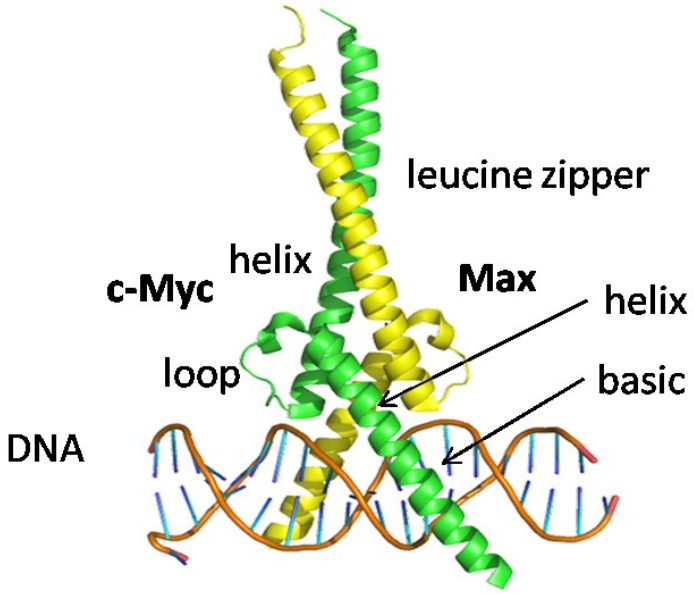

Figure 1. Structure of the c-Myc/Max bHLHZip (basic Helix-Loop-Helix Leucine Zipper) dimerization domains bound to DNA. c-Myc green, Max yellow. PDB: 1NKP.

c-Myc/Max participate in the so-called "Myc network" [18-22] (Figure 2). Opposing the effects of c-Myc/Max heterodimers are four additional closely related members termed the "Mad" family (Mad1, Mxi1, Mad3, and Mad4). Myc network members both positively and negatively regulate gene expression. c-Myc/Max heterodimers bind to the "E-box" of target genes and activate the expression of adjacent genes (+), via the N-terminal transactivation domain (TAD) of c-Myc [16,17]. The TAD recruits the large ATM/PI3-kinase-related protein TRRAP and associated histone acetylases (HAc). The resulting complex is believed to remodel adjacent chromatin, thus facilitating access by other transcription factors [19,21]. Mad/Max heterodimers compete for the same binding sites and repress transcription (-) through the formation of complexes containing the co-repressor mSin3 and associated histone deacetylases (HDAc) [19]. These complexes oppose the effect of c-Myc by de-acetylating chromatin and thus rendering it transcriptionally inactive.

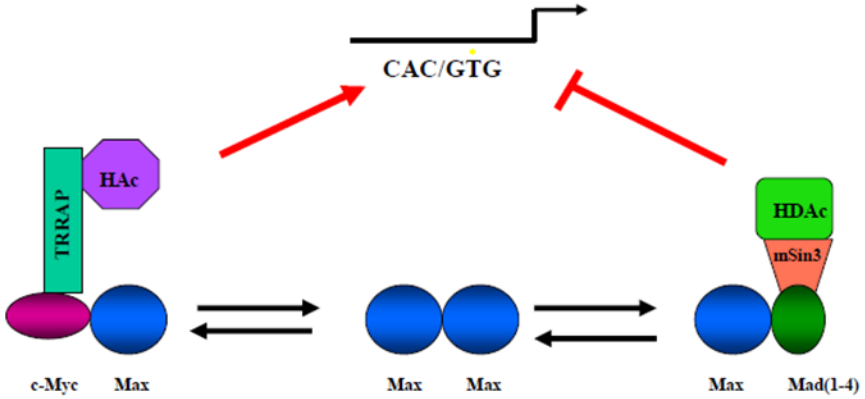

Figure 2. The "Myc network" members both positively and negatively regulate gene expression.

The fact that c-Myc promotes cell proliferation and is over-expressed in many cancer cells [23,24] makes it an attractive anti-cancer target [25]. Since all known functions of c-Myc are dependent on its ability to heterodimerize with Max, small-molecule c-Myc/Max disruptors are pursued as potential anti-cancer drugs. Seven small molecules that prevent c-Myc/Max association and abrogate c-Myc's biological activities have been discovered by Prochownik and collaborators [26,27]. These compounds have been shown to bind independently to one of three distinct sites on c-Myc, and NMR models of these sites bound to a representative compound have been developed [28]. Subsequently, Mustata *et al.* [29] discovered new c-Myc/Max antagonists using a 3D pharmacophore model developed from one parental compound and its analogues. Unfortunately, all compounds discovered thus far presented pre-clinical development issues, such as toxicity, rapid metabolism and lack of activity in cells. An alternative approach to interfere with c-Myc's biological activity was demonstrated by Jiang *et al.* [30], who reported the discovery of agonists of Max-Max homodimers using virtual screening and showed that such agonists reduced cell proliferation, presumably, by reducing the availability of Max monomers to heterodimerize with c-Myc.

2.0 CURRENT APPROACHES AND CHALLENGES IN THE DISCOVERY OF MODULATORS OF PROTEIN-PROTEIN INTERACTIONS

This chapter is based on the review "Discovery of modulators of protein-protein interactions: current approaches and limitations" by L. Meireles and G. Mustata (*Current Topics in Medicinal Chemistry*, 2011, 11:248-257) [31]. It presents a critical review of the general strategies for discovering small-molecule modulators of PPIs, namely: high-throughput screening, fragment-based drug discovery, peptide-based drug discovery, protein secondary structure mimetics, and computer-aided drug discovery. Given the broad scope of the subject, focus is given to the high-level principles underlying each strategy as well as their advantages and disadvantages. For each strategy, at least one example of successful discovery of small-molecule modulator of PPI is taken from recent literature.

2.1 HIGH-THROUGHPUT SCREENING

High-throughput screening (HTS) technologies enable rapid, automatic screening of hundreds of thousands of compounds for therapeutic applications in reasonable timeframes (i.e., weeks). Since the late 1980s, continuous technological advances in fields such as instrumentation, robotics and assays technologies, as well as methodological advances in the areas of assay optimization, control and data analysis have contributed to consolidate HTS as a major drug

discovery paradigm [32,33]. However, the success of HTS in delivering starting points for lead optimization has not been uniformly distributed among different drug target classes; typically, HTS has been more profitable for traditional and well-studied targets (i.e. enzymes and small-molecule receptors [6]) and less so for PPI targets, where the hit rates have been lower [34,35].

One serious limitation of the screening paradigm is that the success or failure of a screening campaign is highly dependent on the compatibility of the molecular library with the target. HTS will fail to deliver useful lead series if the library screened is chemically and/or sterically unsuitable for the target on hands. The fact that HTS is typically more successful when applied to traditional targets, as opposed to PPIs in general, may reflect the bias introduced on current molecular libraries by past medicinal chemistry efforts. It is now gaining acceptance that in HTS compound quantity should not be emphasized over quality. By exploiting prior knowledge about the PPI target, such as its structure, known ligands, and moieties known or predicted to interact with hotspot areas on the protein-protein interface or allosteric binding sites, one may influence the selection of molecules to screen and design PPI-specific molecular libraries to increase the chances of identifying useful hits [36]. However, if the target is unprecedented and no sufficient information is available to build a focused library, then screening a large and chemically diverse library might be more prudent, or another drug discovery strategy could be applied.

HTS is especially useful if the goal is to quickly assess the activity of a large number of compounds in cell-based assays. One of the most promising developments in this area is the so-called *high-content screening* (HCS) [37], an advanced form of cell-based HTS that exploits fluorescence bioimaging technologies to record spatiotemporal data about properties of individual cells, and automatic bioimage analysis to translate the information embedded on cell

images into phenotypes and their changes induced by potential drugs. Several HCS assays [38-40] have been developed aiming at screening for small-molecule modulators of PPIs in living cells. Recently, Colas [41] reviewed the repertoire of assay technologies available for both *in vitro* and cell-based screening to discover inhibitors of PPIs, whereas Fletcher and Hamilton [42] reviewed several successful cases of PPI inhibitors discovered from screening techniques.

One example of proven success using HTS is the discovery of two classes of inhibitors of the interaction between the tumor suppressor protein p53 and its negative regulator MDM2/HDM2 (murine double minute 2 / human double minute 2) [43,44] (Figure 3A).

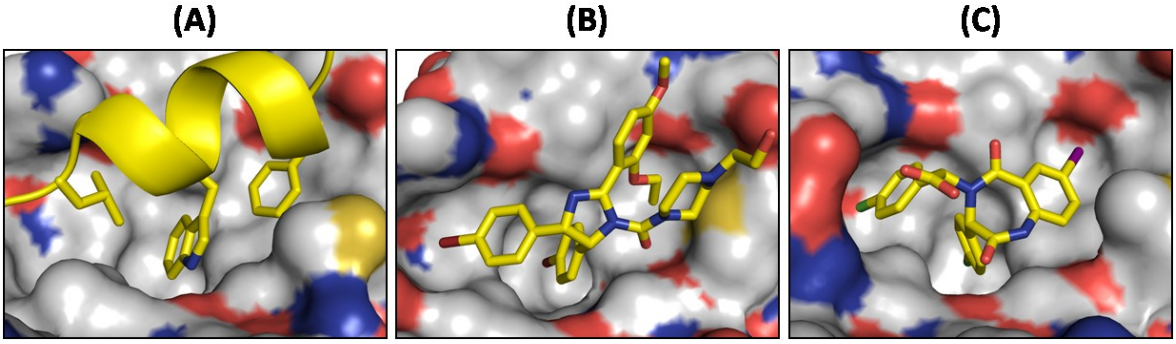

Figure 3. Inhibitors of p53-MDM2 (HDM2). MDM2 is shown in surface representation and ligands in sticks. (A) p53 α-helix bound to MDM2, PDB:1YCR; (B) Nutlin-2 compound (IC50 = 140 nM), PDB:1RV1; (C) benzodiazepinedione derivative (K_d = 80 nM), PDB:1T4E.

Preventing the interaction between p53 and HDM2 with small organic molecules constitutes an attractive strategy for developing anti-cancer drugs [45], since over-expression of HDM2 in many cancer cells impairs p53 tumor suppressor function [46]. To identify p53-MDM2/HDM2 antagonists, researchers at Hoffmann-La Roche [43] screened a diverse library of compounds. Then, a series of cis-imidazoline analogs, named Nutlins, was discovered following optimization of the lead structures. The most potent compound of this series, Nutlin-3, had IC50 = 90 nM. Figure 3B shows the complex of an analog, Nutlin-2, with MDM2 (IC50 = 140 nM). Another class of p53-MDM2/HDM2 antagonists was discovered by researchers at Johnson &

Johnson [44] by employing a miniaturized thermal denaturation assay to screen a library of more than 338,000 compounds for binding to HDM2. Out of 1,216 hits selected for further characterization, 116 compounds were from a benzodiazepinedione series, which yielded a compound with K_d = 80 nM after optimization (Figure 3C).

2.2 FRAGMENT-BASED DRUG DISCOVERY

Fragment-based drug discovery (FBDD) is a relatively new drug discovery paradigm that has been successfully applied to challenging targets, including PPIs [47-55]. In FBDD, a small library of molecular fragments (MW < 250 Da), containing up to a few thousand molecules, is first screened for binding activity against a target of interest. Subsequently, binding fragments are elaborated onto high-affinity compounds by growing a fragment (i.e. fragment evolution), by linking two fragments, or by merging fragments based on a common substructure.

Since molecular fragments bind with low affinity, often in the millimolar range, sensitive biophysical methods are primarily used to detect fragment binding. Nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography (XRC) and surface plasmon resonance (SPR) are typical biophysical methods used for screening fragment libraries [56-58]. While SPR has higher throughput compared to NMR and XRC, it cannot provide information on specific binding modes of molecular fragments, whereas NMR and particularly XRC can reveal such extremely useful pieces of information. Other biophysical methods applicable to fragment screening include mass spectrometry (MS) and isothermal titration calorimetry (ITC). Biochemical functional assays have also been successfully applied [59,60]; however, because fragments are low-affinity binders, it is necessary to use high compound concentrations, which may interfere

with the biochemical assay and result in high false-positive rates. A more specialized approach to fragment screening, called tethering, explores the reversible covalent bond formation, through thiol-disulfide exchange, between a cysteine residue near the binding site and disulfide-contained fragments [61]. In tethering, the identification of the covalently bound fragments is performed using mass spectrometry.

Screening molecular fragments often results in higher hit rates than conventional HTS [51,54]. The chemical space of fragments is orders of magnitude smaller than the chemical space of higher molecular-weight compounds, such as drug-like compounds, and therefore the gap between the chemical space and what is effectively screened is much smaller for fragments. From a different perspective, one may think of the chemical space of drug-like compounds as being formed by combinations of small fragments; screening fragments thus avoids the combinatorial explosion of higher molecular-weight compounds, probing more chemical space with fewer molecules [48,52,54]. This efficient sampling of the chemical space is thought to contribute to higher hit rates.

Also contributing to higher hit rates is the fact that fragments tend to be promiscuous binders [48,52,54]. That is, due to their smaller size and lower molecular complexity, fragments are more likely to optimally accommodate themselves on different target binding sites without causing clashes or making unfavorable interactions [62]. Nevertheless, when the binding fragments are combined into larger and high-affinity compounds, their molecular complexity increases and so does their specificity. Therefore, fragment promiscuity may not necessarily constitute a complication and, on the contrary, it may contribute to higher hit rates [43,49,51]. Contrasting to that is the promiscuous inhibitors discovered by Shoichet and co-workers [63] which frequently appear as false-positives in HTS campaigns. Those promiscuous inhibitors are

uninteresting to drug discovery because they act non-specifically through the formation of large molecular aggregates.

Despite the fact that fragments are low-affinity binders, they often bind with high efficiency, with most of their atoms directly engaging in strong interactions with atoms on the target binding site [48,52,54]. Ligand efficiency, usually defined as the binding free energy normalized by the number of heavy atoms is often used as a metric for prioritizing fragment hits for lead development and optimization [64,65].

The main challenge of FBDD is to transform fragment hits into leads and drug candidates. Knowledge of the fragments' binding modes on the protein target allows the hits to be optimized using rational, structure-based design of compounds. While structural knowledge considerably facilitates fragment growing and optimization, it is possible to optimize fragment hits in the absence of structural information; for example, by screening a second-generation molecular library consisting of random extensions of the fragment hits. Although a certain degree of serendipity is required, this procedure is useful not only to those targets that are elusive to structure determination but also to targets whose binding site flexibility is difficult to model (see e.g. [13]).

A successful example of FBDD applied to PPIs is the targeting of anti-apoptotic members of the Bcl-2 family, namely Bcl-x_L, Bcl-w and Bcl-2 proteins, by researchers at Abbott Laboratories [66]. Over-expression of these proteins in many cancer types correlates with tumorigenesis and resistance to chemotherapy, presumably by binding and sequestering pro-apoptotic members of the Bcl-2 family [67,68], such as the BH3 α-helix only proteins, Bad and Bid. Abbott carried out a large fragment screening using 2D-NMR to discover fragments that bind to the hydrophobic BH3-binding groove of Bcl-x_L [69]. Two fragment hits, **1** and **2** (Figure

4), were identified and shown to bind to distinct but proximal sites [69]. NMR-based structural studies guided the linking of the fragments as well as subsequent optimization, which led to compound **3** [69]. Compound **3** was then modified to remove binding to human serum albumin [70] and also to improve binding to Bcl-2 [71], resulting in compound **ABT-737**, with high affinity (K_d < 1 nM) to Bcl-x_L, Bcl-2 and Bcl-w proteins. Finally, compound **ABT-263**, an analog of **ABT-737**, was designed to improve oral bioavailability [72]. **ABT-263** is currently undergoing phase II clinical trials for the treatment of small-cell lung cancer, and other malignancies [73]. Figure 5 illustrates the crystal complex structure of **ABT-737** with Bcl-x_L.

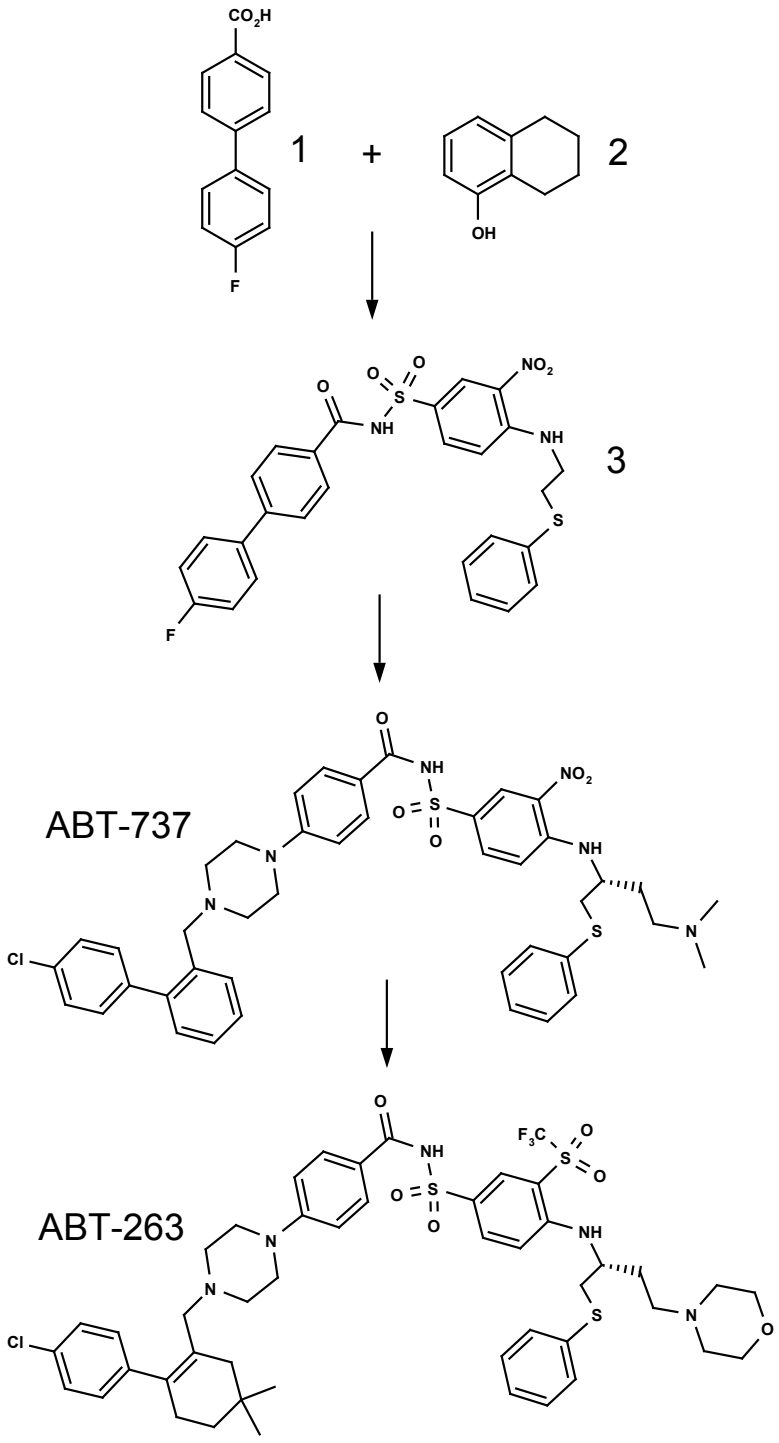

Figure 4. From fragments hits to clinical trial candidate ABT-263, antagonist of Bcl-x_L, Bcl-2 and Bcl-w proteins.

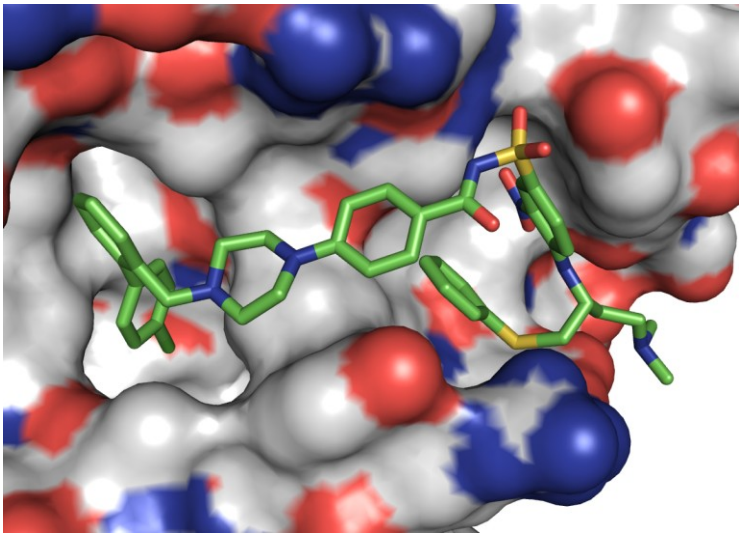

Figure 5. Crystal structure of Bcl-x_L (surface) in complex with ABT-737 (sticks), PDB: 2YXJ.

2.3 PEPTIDE-BASED DRUG DISCOVERY

Peptides are powerful tools in drug discovery serving several roles, including their use as probes in target validation and chemical genetics studies, as surrogate ligands in HTS competitive binding assays, and as intermediate leads for developing non-peptidic small molecule drugs [74]. Furthermore, peptides themselves have been developed into therapeutic agents [75,76]. All these applications are facilitated by the availability of well-established *in vitro* protein evolution technologies that can discover and optimize peptides for virtually any protein target [77-79]. These technologies rely on principles of diversification, selection and amplification to mimic nature's evolution *in vitro* and optimize peptides for binding to a target. Among these technologies, phage display has been the most commonly used to optimize peptides [77-80].

The directed evolution of peptides toward targeting protein-protein interfaces often delivers peptides that mimic the binding epitopes of one of the protein partners, albeit using

much less scaffold [81,82]. In some cases, it delivers peptides bearing novel binding epitopes that are unexpected from the analysis of the protein structure, and/or peptides with higher binding affinities than the target's natural binding partner. The task of targeting PPIs with peptides is considerably facilitated by the availability of *in vitro* protein evolution methods. Perhaps contributing to that is the ability of peptides to assume extended conformations that better match the typically large protein-protein interfaces [82].

While peptides can be made to target proteins with high specificity and affinity, they have generally poor pharmacokinetics properties, such as metabolic instability, low oral bioavailability and low cell membrane penetration. These properties present considerable obstacles to their widespread use as therapeutic agents. In order to overcome these limitations, several peptidomimetic strategies have been developed. The general approach is to introduce chemical modifications to the peptide so as to reduce its peptidic character [83-85] and improve its metabolic stability and delivery [86-88]. Despite recent progress in this direction, the evolution of a peptide lead into a peptide drug remains a challenging problem. Nevertheless, peptides are still useful as intermediate leads on which to base the design of non-peptidic small-molecule compounds that mimic the essential binding features of the peptide.

The following example illustrates the use of a peptide lead to develop small-molecule antagonists of PPIs. The second mitochondria-derived activator of caspases (SMAC) is an endogenous pro-apoptotic protein that antagonizes several IAP (inhibitor of apoptosis) proteins, including XIAP (X-linked IAP), cIAP-1 and cIAP-2 [89]. SMAC promotes programmed cell death by binding its N-terminal tetrapeptide motif Ala-Val-Pro-Ile to IAP proteins [90,91], thereby preventing IAPs from inactivating caspases [92], which are required for apoptosis. Therefore, SMAC mimetic compounds have the potential to serve as anti-cancer drugs in cancer

cells that evade apoptosis by over-expressing IAPs and under-expressing SMAC. Figure 6 shows a few SMAC-mimetic XIAP antagonists that have been designed[93,94] based on the structure of the peptide motif AVPI bound to the BIR3 domain of XIAP (Figure 7A). The structure of compounds **7-8** modeled on top of the peptide shows that the essential binding features of the peptide are retained by the compounds (Figure 7B-C) [94]. Recently, a SMAC mimetic compound named AT-406 [95] (Ascenta Therapeutics; chemical structure not disclosed) has entered Phase I clinical trials for the treatment of advanced solid tumors and lymphomas.

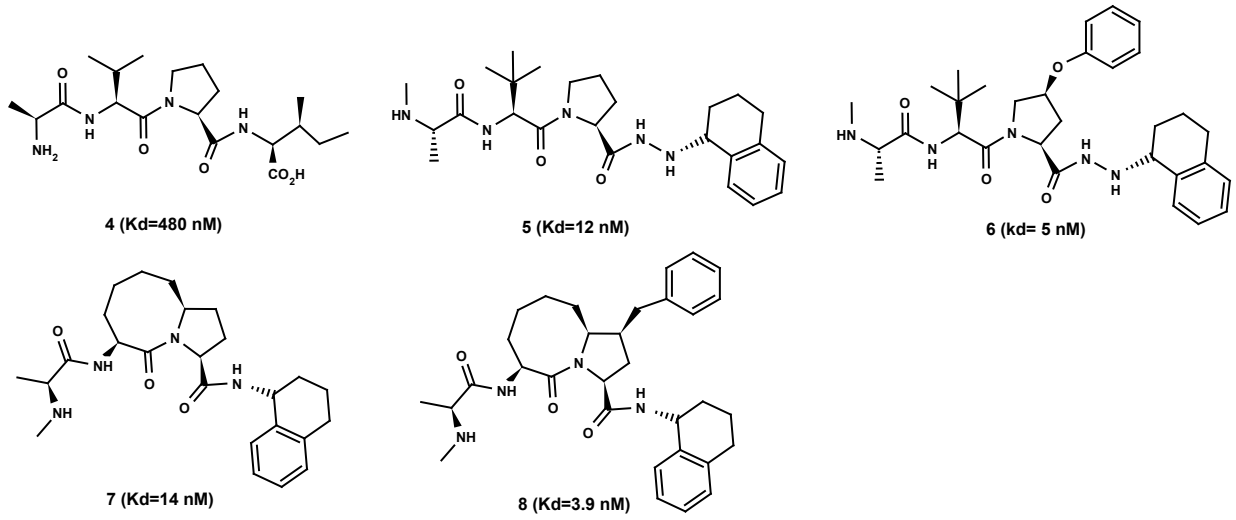

Figure 6. SMAC mimetic compounds and their binding affinity to XIAP. Compound 4 is the tetrapeptide AVPI.

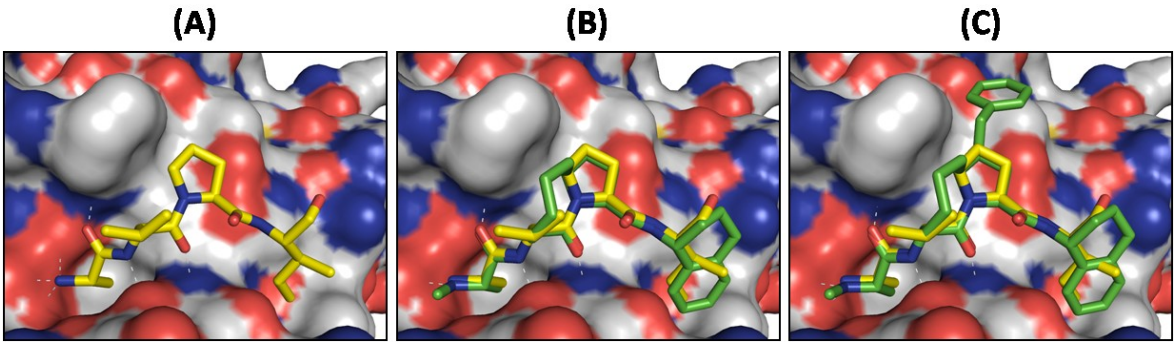

Figure 7. XIAP antagonists. (A) the tetrapeptide AVPI bound to the BIR3 domain of XIAP, PDB:1G73; (B) and (C) Compounds 7 and 8, respectively modeled on top of the peptide, mimic the binding features of the peptide.

Similarly to the previous example, many PPIs in signal transduction pathways occur via a short peptide sequence, about 4-8 amino acids long, known as *linear peptides*. In signal transduction, linear peptides are often recognized by modular or adaptor domains such as the members of the following families: PDZ, SH2, SH3, PTB and WW. PPIs mediated by linear peptides are natural candidates for peptide-based drug discovery, since the linear peptide by itself can act as an antagonist lead, by competing with its parental protein for the same binding partner and serving as a starting point for medicinal chemistry. Selective peptide modulators of PPIs in intracellular signaling are not only useful probes to decoding the function of interacting proteins in signaling cascades, but they are also potential leads for developing non-peptidic drugs targeting those interactions [96]. In particular, members of the PDZ and the SH2 domain families have been pursued as therapeutic drug targets and several peptidic and non-peptidic small molecules have been discovered to prevent protein-protein associations mediated by these two domains [97,98].

An interesting example of peptides designed to target an extracellular PPI is demonstrated by a series of studies that attempted to reproduce with a short peptide the agonist activity that 166-residue hematopoietic growth hormone, erythropoietin (Epo), induces on to its

receptor, Epo-R. Binding of Epo activates Epo-R via reorientation of two monomeric chains of Epo-R, thereby activating two Janus kinases 2 (JAK2) in the intracellular region of Epo-R [99]. In turn, the kinases initiate a cascade of events that lead to red blood cell proliferation. Using a random phage display peptide library, Wrighton and colleagues [100,101] discovered a 20-residue cyclic peptide that activated Epo-R, although with less efficiency than the natural ligand Epo. The same group was subsequently able to reduce the peptide to 13 residues, while keeping its agonist activity, via a series of truncated peptides and mutational studies [102]. However, one of the peptide analogs was later found to be an antagonist of Epo-R function [103]. Interestingly, other examples of structurally related agonist and antagonists of peptide receptors can be found in the literature [104]. Analysis of the crystal structures of the extracellular binding domains of Epo-R in complex with each of the two peptides (agonist and antagonist) revealed that the peptides induced different domain orientations [103]. This demonstrates that it is not sufficient to target the receptor binding site of a natural agonist (i.e., Epo) for eliciting agonist activity, but the molecule must also be able to induce specific conformational changes [103,105]. A potent peptide Epo-R agonist, called Hematide^{TM} (Affymax, Palo Alto, CA), is currently in Phase III clinical trials for the treatment of anemia associated with chronic renal failure [106].

2.4 PROTEIN SECONDARY STRUCTURE MIMETIC SCAFFOLDS

α-helices, β-sheets, and β-turns are basic structural elements of proteins frequently involved in mediating protein interactions. The development of small-molecule scaffolds displaying side-chain functionalities with similar spatial arrangement as those presented by protein secondary

structure elements has great potential to deliver compounds that mimic protein-protein recognition sites. The basic assumption is that pre-organization of interacting side-chains presented by semi-rigid scaffolds enhances the binding affinity of the ligand [107]. In the following, we review recent efforts toward the development of α-helix mimetic scaffolds. For a detailed review of protein secondary structure mimetic scaffolds, including β-strand and β-turn mimetic scaffolds, see Marshall et al. [107] and Hershberger et al. [108].

Many protein interactions involve the projection of amino acid side-chains from an α-helix of one protein toward small pockets on the surface of another protein. This mode of interaction is often observed, for example, in intrinsically disordered domains which undergo disorder to helix transition upon binding. Such helices are usually amphiphilic α-helices with hydrophobic residues lying in one face of the helix at amino acid positions i, i+4 and i+7, approximately, facing the binding site, and hydrophilic residues on the other face of the helix, exposed to solvent. Because this is a recurring pattern encountered in many protein interactions, several research groups have recently attempted to design small-molecule scaffolds whose side-chains are presented with similar distance and angular constraints as the side-chains i, i+4 and i+7 of an ideal α-helix (see Figure 8A).

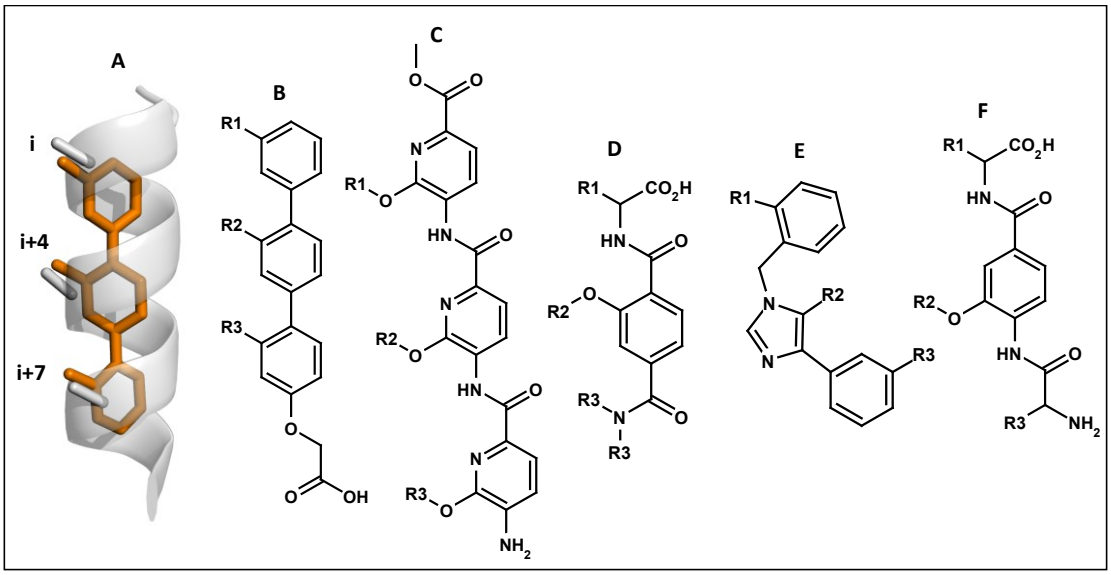

Figure 8. α-helix mimetic scaffolds mimic the presentation of side-chains from one face of an α-helix. (A) terphenyl scaffold superimposed on an ideal α-helix; (B) terphenyl scaffold [109]; (C) tripyridylamide scaffold [110]; (D) terephthalamide scaffold [111]; (E) tri-substituted imidazole scaffold [112]; (F) Boger et al. scaffold [113].

Hamilton and co-workers [109] first proposed a tri-substituted terphenyl scaffold to mimic the presentation of side-chains from one face of an α-helix (Figure 8A-B). The scaffold was applied to mimic the α-helix of the smooth muscle myosin light chain kinase (smMLCK) which binds calmodulin (CaM). Three side-chains of this α-helix bind to CaM: Trp800, Thr803, and Val807. By attaching analogs of these side-chains (respectively, 1-naphthyl, ethyl and isopropyl) into the terphenyl scaffold, a potent small-molecule antagonist of the smMLCk-CaM interaction (IC50 9 nM) was designed. Subsequently, Hamilton and co-workers [114] applied the same terphenyl scaffold to antagonize the interaction between Bcl-x_L and the pro-apoptotic proteins Bak and Bad. In this case, the helical BH3 binding domain of Bak served as the template for designing several Bcl-x_L inhibitors, of which the most potent had K_d value of 114nM.

One problem with the terphenyl scaffold is that it is highly hydrophobic and thus poorly soluble in water. This problem was ameliorated with the introduction of a similar but more hydrophilic scaffold based on an oligoamide foldamer (tripyridylamide) [110] (Figure 8C). This scaffold has the additional advantage of pre-orienting the side-chains through the formation of intra-molecular hydrogen bonds. A significant improvement in water solubility was later accomplished with a new proposal, the terephthalamide scaffold [111] (Figure 8D), which significantly simplified the terphenyl scaffold by replacing the flanking phenyl rings with carboxamide groups. The terephthalamide helix mimetic scaffold was also applied to disrupt the interaction between Bcl-x_L and Bak, delivering several compounds with dissociation constants in the low micromolar range [111]. Dömling and co-workers [112] proposed an alternative α-helix mimetic scaffold based on a tri-substituted imidazole backbone (Figure 8E) accessible by multicomponent reaction chemistry. Several derivatives of this scaffold showed micromolar potency toward inhibiting the interaction between Bcl-w and Bak. More recently, Boger and co-workers [113] synthesized a library of 8,000 helix mimetic compounds based on the scaffold shown in Figure 8F. The library was screened for antagonists of the MDM2-p53 interaction and a compound that mimicked the p53 α-helix was identified.

2.5 COMPUTER-AIDED DRUG DISCOVERY

Drug discovery is a time-consuming, multibillion dollar endeavor with high risks involved. In order to improve productivity in drug discovery and reduce research costs, computational methods are indispensable. This section discusses the two typical scenarios in which computer-

aided drug design can significantly accelerate drug discovery. These scenarios are structure-based drug design and ligand-based drug design.

2.5.1 Structure-based drug design

Structure-based drug design (SBDD) relies on the availability of one or more experimentally determined protein structure to generate hypotheses about suitable small-molecule ligands [115-117]. However, the experimental determination of the target protein structure, a pre-requisite for SBDD, is not a trivial task. For some targets, such as membrane proteins and intrinsically disordered proteins, structural determination may be very difficult to achieve.

Virtual screening is the most prominent example of SBDD where the goal is to rank-order compounds according to their shape and chemical complementarities to the target's binding site. Virtual screening involves docking and scoring. Docking is the computational method for searching the space of ligand poses on the binding site. Scoring associates a weight to each pose as a measure of its fit. Several scoring functions have been developed; however, they generally suffer from inaccurate treatment of solvation and entropic effects. Nevertheless, virtual screening methods continue to improve and have already demonstrated practical utility in many drug discovery campaigns. A successful example of virtual screening for PPIs is the identification of stabilizers of the Max homodimer, an indirect way of interfering with the formation of the c-Myc/Max heterodimer [30].

De novo design of ligands is another example of the SBDD approach whereby a ligand is built "from scratch" within the target binding site. The main drawback of this approach is the lack of synthetic feasibility of the resulting compounds. One of the major challenges of structure-based approaches has been the modeling of protein flexibility. Structure-based methods are

currently unable to accurately predict the binding modes of ligands to proteins that undergo large structural rearrangements upon ligand binding [118], a fact that limits the applicability of structure-based methods. Recent advances in modeling large conformational changes of proteins observed upon binding have been made recently using Elastic Network Models. We will revisit this topic in Chapter 3.0 .

2.5.2 Ligand-based drug design

Ligand-based drug design (LBDD) exploits the information about known active compounds (and possibly inactive compounds as well) to discover new actives [119]. Ligand-based approaches rely on the central similarity-property principle, which states that similar molecules should exhibit similar properties [120]. Hence, the activity prediction of a compound or a set of compounds is made based on the similarity or distance to a set of reference ligands with known bioactivity to a protein target [121]. Different types of two- and three-dimensional molecular descriptors and substructures in combination with a variety of classification schemes, such as decision trees, Bayesian statistics, neural networks or other machine-learning methods have been used for this purpose.

Pharmacophore-based virtual screening can be considered to be at the intersection between structure-based and ligand-based approaches, as either the protein structure or known ligands can be used as references to build the models (respectively referred to as, receptor-based pharmacophore and ligand-based pharmacophore). One advantage of similarity searching over pharmacophore-based search is that it does not require a set of structurally unrelated compounds of similar biological activity to derive a model. When using similarity searching, even one active molecule can be used to search a database for related compounds. This similarity-based virtual

screening has proven very convenient, as it is computationally inexpensive and requires little information [122]. It mostly uses 2D descriptors, also called topological descriptors, which are derived from the connectivity table of the molecule and take into account distances among atoms in terms of number of bonds in the shortest path between them. The most commonly used descriptors are topological fingerprints [123], which encode the presence or absence of substructural fragments in molecules in a binary fingerprint, without taking into account the number of occurrences of the feature. These fingerprints can be pre-calculated and compared, usually by means of Tanimoto distance, in a very fast and efficient manner to any reference set. In terms of the novelty of the leads discovered, however, pharmacophore search provides more scaffold hopping capability than fingerprint-based methods.

One example of a ligand-based pharmacophore modeling for PPIs is the development of a 3-dimensional pharmacophore model using a set of known inhibitors of c-Myc/Max heterodimer formation [29]. This was the first report of a pharmacophore model to provide a hypothetical picture of the main chemical features responsible for the activity of c-Myc/Max heterodimer disruptors. The model successfully identified a set of structurally diverse compounds that showed affinities in the μM range and growth inhibitory activity against c-Myc-overexpressing cells [29].

2.6 FINAL REMARKS

Unfortunately, there is no single best drug discovery paradigm capable of tackling all drug targets. It is thus important to understand the strengths and weaknesses of each approach, and realize that, in fact, they are complementary approaches. The synergism created by combining

approaches, e.g., combining experimental with computational approaches in particular, may be the key to increase productivity in this field. New classes of drug targets require new ways of thinking, and as scientists experience the transition from enzymes and small-molecule receptors to PPI targets the field will further develop in the coming years.

3.0 MODELING PROTEIN FLEXIBILITY COUPLED TO LIGAND BINDING USING ELASTIC NETWORK MODELS

This chapter is based on the review "Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins" by L. Meireles, M. Gur, A. Bakan and I. Bahar (in press in *Protein Science*) [124].

Modeling protein flexibility is a challenge task owing to the enormous size and complexity of proteins' conformational space - which preclude exhaustive sampling - and the prohibitively large computational times required to perform fully atomistic molecular dynamics simulations up to the timescales of biological significance. Moreover, the complexity of the problem is compounded by inaccuracies of scoring functions and force fields.

When dealing with complex problems, robust simplifications hold great promise. Along those lines, coarse-grained normal mode analyses (NMAs) have been used to model large-scale collective motions of proteins. In particular the NMA of unbound proteins using Elastic Network Models (ENMs) proved to predict soft modes of reconfiguration in accord with conformational changes experimentally observed upon ligand binding [125,126], implying that proteins' intrinsic dynamics plays a major role in defining the bound conformations. We present here a summary of the foundations of ENMs-based NMA for modeling the collective motions of proteins, and several recent applications showing how they have been effectively adopted for modeling the conformational changes of proteins observed upon ligand binding. Most importantly, these

studies help improve our understanding of the molecular basis of observed recognition and binding events, and stipulate the importance of the energy landscape, or the *soft paths/modes of motion* away from the native state energy minimum which predominantly define the bound conformers being selected.

3.1 PROTEIN FLEXIBILITY IN MOLECULAR RECOGNITION AND BINDING

Several molecular recognition models have been proposed with different perspectives for the role of protein flexibility in molecular recognition and binding. Fischer's *lock-and-key* model [127] postulates that enzymes and their substrates are rigid, complementary-shaped bodies that fit each other as a lock and key. Although the importance of shape (and chemical) complementarity to binding is widely acknowledged nowadays, the lock-and-key model neglects protein flexibility by treating proteins as rigid molecules. Since this model has been proposed (more than a century ago), advances in experimental technologies have improved our understanding about proteins, and it is now clear that protein flexibility varies over a spectrum ranging from fairly rigid globular proteins to intrinsically disordered proteins - characterized by the absence of stable secondary structures. Most proteins, however, are comprised of combinations of rigid (e.g., hydrophobic core) and flexible substructures (e.g., loops and hinge residues), evolved to collectively confer the protein the dynamics required to perform its function.

To explain the structural changes of proteins observed upon ligand binding, Koshland proposed the *induced-fit* model [128], whereby the ligand drives the conformational changes of the protein structure to optimize protein-ligand interactions (Figure 9A). Contrasting to this view, the *conformational selection* model proposed by Monod, Wyman, and Changeux (MWC model)

[129] postulates that (a) an ensemble of protein conformations pre-exists in dynamical equilibrium prior to ligand binding; and (b) the ligand binds to, and stabilizes, one such conformation, shifting the equilibrium towards the bound state (Figure 9B) – see Boehr *et al.* [130] for a review of recent experimental evidence in support of conformational selection.

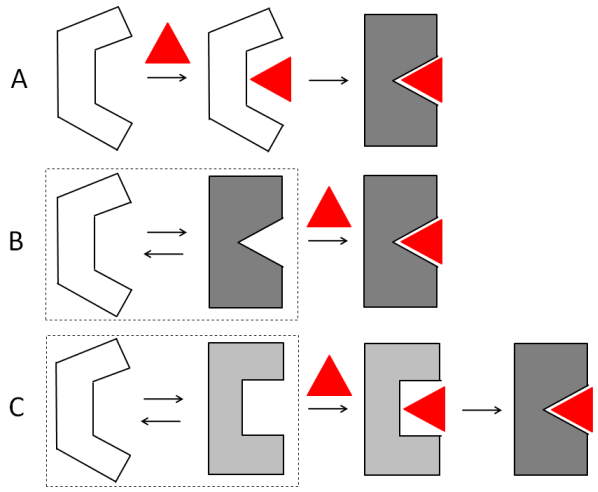

Figure 9. Schematic diagram of models for ligand binding. The protein is shown in white/gray, and the ligand in red. (A) Induced-fit; (B) Conformational selection; (C) Conformational selection followed by induced-fit. See text for details. In parts B and C the protein is originally in dynamic equilibrium with an ensemble of fluctuating conformers (boxes at the left), represented here by two conformers for simplicity**.**

Despite fundamental differences between the induced-fit and conformational selection models, it is likely that both mechanisms may co-occur in binding: conformational selection might play a dominant role in defining the global/large conformational changes (intrinsic to the protein) that are exploited for ligand binding, especially at the early stage of binding; and induced-fit would play a role in inducing local and ligand-specific changes in the vicinity of the ligand, such as side-chain flips and loop rearrangements, so as to further optimize the interactions (Figure 9C).

3.2 ELASTIC NETWORK MODELS: THEORY AND METHODS

Elastic Network Models (ENMs) have been widely used in recent years for investigating the cooperative motions that biomolecular systems tend to undergo under equilibrium conditions. The basic assumption in ENMs is that the dynamics of a protein is uniquely defined by its contact topology, represented as a network of nodes and springs.

ENMs were originally inspired by the work of Tirion [131], who demonstrated that the global modes obtained by Normal Mode Analysis (NMA) with a detailed force field are almost identically reproduced by adopting a single-parameter harmonic potential between all atom-pairs within a short interaction range (≈ 5.0 Å). Several studies have confirmed since then that the global modes of biomolecules are robustly defined by the overall shape, or inter-residue contact distribution, irrespective of the detailed structure and energetic [126,131-135]. Most importantly, these modes have been shown in many studies to be relevant to functional motions, hence the significance of identifying them by computational methods.

Two most widely used ENMs are the Gaussian Network Model (GNM) [136,137] and the Anisotropic Network Model (ANM) [138-140], where the node positions are identified by the coordinates of α-carbons known from experiments, and all residue pairs whose α-carbons are located within a cutoff distance r_c are connected by a spring of uniform force constant γ. The respective underlying potentials in these two models are:

Equation 1

$$V_{\mathrm{GNM}} = -\frac{\gamma}{2}\left[\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(\mathbf{R}_{ij}-\mathbf{R}_{ij}^{0}\right)\cdot\left(\mathbf{R}_{ij}-\mathbf{R}_{ij}^{0}\right)\Gamma_{ij}\right]$$

in accord with the statistical mechanical theory of polymer networks[141], and

Equation 2

$$V_{\mathrm{ANM}} = -\frac{\gamma}{2}\left[\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(R_{ij}-R_{ij}^{0}\right)^{2}\Gamma_{ij}\right]$$

where $\boldsymbol{R_{ij}^{0}}$ and $\boldsymbol{R_{ij}}$ are the equilibrium and instantaneous distance vectors between residues *i* and *j*, R_{ij}^{0} and R_{ij} are their magnitudes, and Γ_{ij} is the ij^{th} element of the Kirchhoff matrix $\boldsymbol{\Gamma}$, the off-diagonal elements of which are defined as $\Gamma_{ij} = -1$ if $R_{ij}^{0} < r_c$ and zero otherwise, and the diagonal elements are evaluated from the summation $\Gamma_{ii} = -\sum_j \Gamma_{ij}$ over all off-diagonal elements in the i^{th} row (or column). As such, the GNM potential includes contributions from both distance and orientation changes, whereas the ANM potential is exclusively based on distance changes.

A coarse-grained (single-node-per-residue) harmonic potential for all residue pairs was first proposed by Hinsen [142,143], the difference from the ANM being the adoption of a distance-dependent force constant (thereby eliminating the parameter r_c). Using a uniform force constant for each pair of residues within a given cutoff treats both strong/specific (covalent bonds between consecutive residues or hydrogen bonds within helices and sheets) and weak/nonspecific (hydrophobic contacts between side-chains) interactions identically. Variations of ANM using more elaborate force constants have been shown to slightly improve the accuracy of predictions [144,145]. Examples include force constants weighted by an exponential decay function [142], negative exponents [144,146], and another form combining negative exponent and contact area between residues [147]. Recently, using an entropy maximization method, Lezon and Bahar [148] further investigated the determinants of structural dynamics. To this aim, they optimized the GNM force constants based on the covariance matrices derived from NMR ensembles. Their study suggests that the next level of refinement in ENMs could be to

incorporate secondary structure-dependent force constants, i.e. stiffer springs for residues forming hydrogen bonds in a helix or a sheet.

The spectrum of ANM normal modes is obtained by NMA using the potential V_{ANM} (Equation 2), which is simply found by eigenvalue decomposition of the Hessian **H**, the matrix of the second derivatives of V_{ANM} with respect to residue position. The eigenvalue decomposition of **H** yields $3N$-6 non-zero eigenvalues and corresponding eigenvectors. The k^{th} eigenvector, $\boldsymbol{u}_k = (u_{x1}\ u_{y1}\ u_{z1}\ \ldots\ u_{zN})_k^{\mathrm{T}}$ describes the normalized displacements of the N residues in the x-, y- and z- directions as driven by mode k, and the corresponding eigenvalue, λ_k, provides a measure of the frequency (squared) of that mode, or the statistical weight (equal to $1/\lambda_k$) of square-displacement along this mode. The change $\boldsymbol{\Delta R}|_k$ in the configuration $\boldsymbol{R}^0 = [x_1^0\ y_1^0\ z_1^0\ x_2^0\ \ldots\ z_N^0]^{\mathrm{T}}$ caused by the fluctuation along ANM mode k is conveniently expressed as [149]:

Equation 3

$$\boldsymbol{\Delta R}|_k = \pm s\,(k_B\mathrm{T}/\lambda_k)^{1/2}\,\boldsymbol{u}_k$$

where $\pm s$ *i*s a variable uniformly scaling the size of the fluctuation along mode k, T is the absolute temperature, and k_B is the Boltzmann constant.

3.3 COMPARISON OF COMPUTATIONAL PREDICTIONS WITH THE PRINCIPAL MODES OF BINDING OBSERVED IN EXPERIMENTS

3.3.1 Metrics for comparison

When two structures (A and B) are available for a given protein, a metric of structural change is the 3*N*-dimensional *deformation vector* $\boldsymbol{d} = \mathbf{R}_A - \mathbf{R}_B$ obtained from the difference between the coordinates of all α-carbons after optimal superimposition of the structures to eliminate rigid-body translational and rotational differences. The correlation cosine between $\boldsymbol{d}$ and ANM mode $\boldsymbol{u}_k$ provides a measure of the level of agreement between experiments and theory. Of interest is the correlation with soft/low-frequency modes (e.g., k = 1-3) to assess whether the experimentally observed (usually functional) changes concur with the 'easiest' reconfigurations the structure intrinsically tends to undergo. As discussed in Section 3.3.2, this has been the case in many applications, suggesting that structures have evolved to favor soft modes that are being exploited during functional changes in conformation.

Previous work has shown that structural ensembles can be advantageously analyzed to extract the principal modes of structural variations, which, in turn, may be compared to ANM soft modes [125]. This is achieved by a principal component analysis (PCA) of the *3Nx3N* covariance matrix, $\boldsymbol{C}$, constructed for the ensemble of structures (e.g., m of them, where $m \leq 3N$ usually) resolved for the studied protein, after optimal iterative superposition of these structures [125]. Eigenvalue decomposition of $\boldsymbol{C}$ as $\boldsymbol{C} = \sum_{i=1}^{m} \sigma_i \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{T}}$, yields the principal components $\boldsymbol{p}_i$ (eigenvectors) and corresponding eigenvalues σ_i. σ_1 corresponds to the largest variance and $\boldsymbol{p}_1$ describes the most dominant structural change. The fractional contribution of mode $\boldsymbol{p}_i$ to

experimentally observed heterogeneity/variance of structures is $\sigma_i / \Sigma_k \sigma_k$ where the summation is performed over all PCA modes. The top three modes usually account for more than 50% of the heterogeneity observed in X-ray crystallographic structures. They also tend to be highly collective in terms of the distribution of the motion among the residues, although in some cases, they may only describe localized (e.g. loop) motions.

The *overlap* between the k^{th} ANM mode and the i^{th} PCA mode is given by the correlation cosine $O_{ik} = \boldsymbol{p}_i . \boldsymbol{u}_k$ [140]. The *cumulative overlap* is used to measure how well a subset of low frequency ANM modes (e.g. *J* of them) predicts a PCA mode *i* and is defined as [150]:

Equation 4

$$CO_i^J = \left[\sum_{k=1}^{J} (O_{ik})^2\right]^{1/2} = \left[\sum_{j=1}^{J} (\boldsymbol{p}_i \cdot \boldsymbol{u}_k)^2\right]^{1/2}$$

Note that $CO_i^J = 1$ for $J = 3N\text{-}6$, i.e., the *3N-6* ANM eigenvectors form a complete set of orthonormal basis vectors.

3.3.2 Comparison of anisotropic network model soft modes with the structural changes experimentally observed between bound and unbound forms

Several studies support the predictive capacity of ENMs and provide a basis for their application to modeling the conformational changes of proteins associated with complex formation. For example, Tobi and Bahar [151] showed for protein-protein complexes with known structures in the *apo* and *holo* forms (LIR-1/HLA-A2, Actin/DNase I, Cdk2/cyclin and Cdk6/$p16^{INK4a}$) that there is a good agreement between the experimentally observed structural changes (between *apo* and *holo* forms) and the collective motions predicted by the ANM for the *apo* structure. For each

case, a single low-frequency ANM mode could be identified with high correlation with the experimentally observed principal structural change $\boldsymbol{p_1}$ (also designated as PC1). Similar results for antigen-antibody complexes were reported by Keskin [152]. Bakan and Bahar [125] further showed that the good agreement was not restricted to PPIs, but protein-small molecule interactions as well, as exemplified by three enzymes (HIV-1 reverse transcriptase (RT), p38 MAP kinase, and Cdk2) for which sufficiently large ensembles of structures in liganded and unliganded forms were available. Notably, similar results were obtained [125] for sufficiently broad ensembles of structures resolved by NMR (residual dipolar coupling measurements) and proteins such as ubiquitin [153] and calmodulin [154]. Figure 10 illustrates the close correspondence between the experimental conformational space and the ANM predictions for the aforementioned enzymes and for calmodulin. Correlations in the range 0.84 - 0.99 are observed. In a similar study involving HIV-1 protease, Yang et al. [150] reported close similarity between the motions predicted by ENMs and those calculated by PCA of a large set of X-ray structures, PCA of NMR ensemble, and PCA of MD simulation snapshots (a.k.a. essential dynamics analysis).

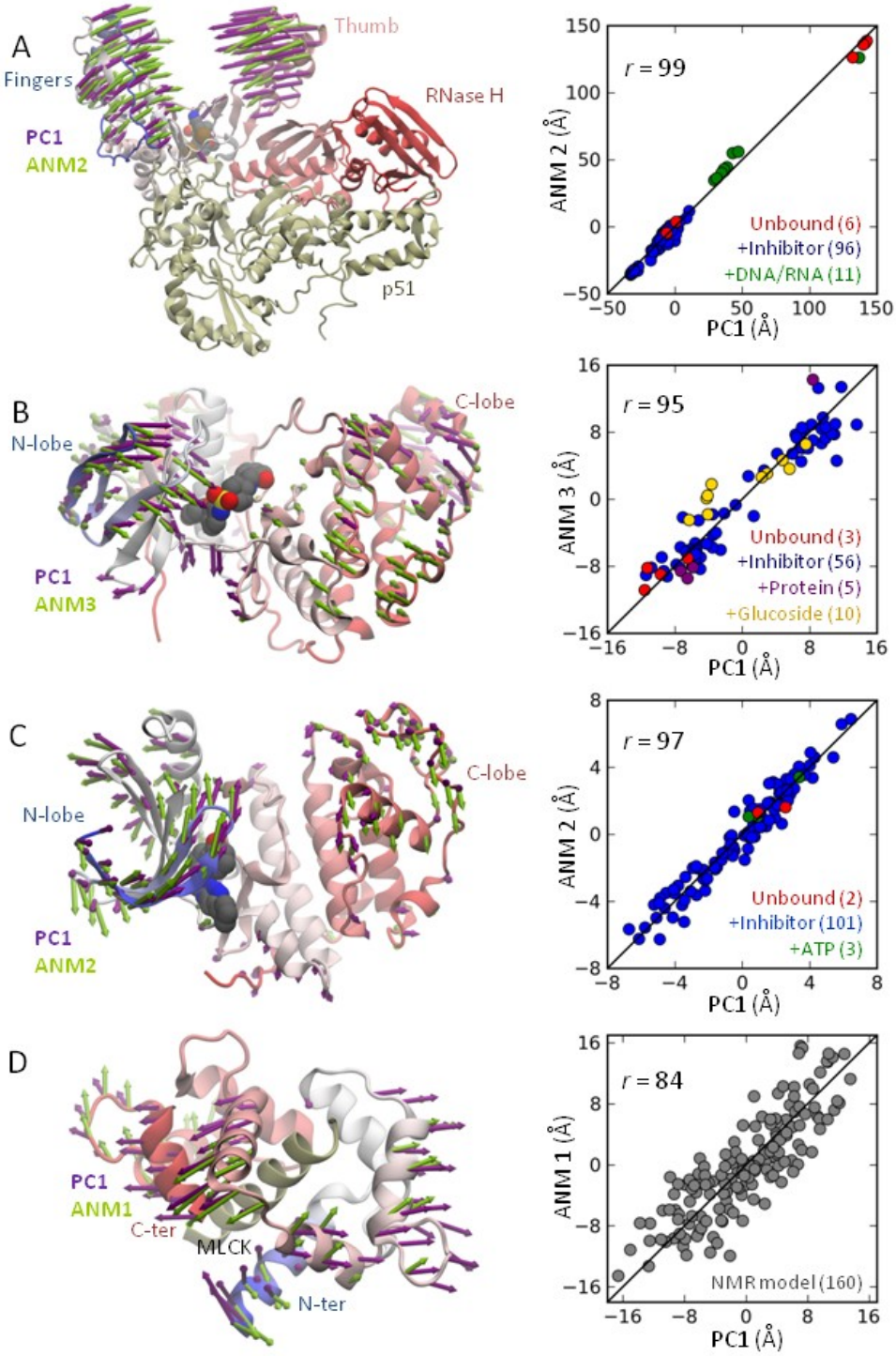

Figure 10. Comparison of principal changes in structure observed in experiments and predicted by the ANM. Results are displayed for (A) HIV-1 RT, (B) p38 MAP kinase (p38), (C) Cdk2, and (D) calmodulin (CaM) complexed with myosin light chain kinase (MLCK) datasets. Experimental datasets (113 HIV- RT X-ray structures, 74 p38 structures, 106 Cdk2 structures, and 160 CaM-MLCK NMR models) were subjected to PCA to obtain the

dominant changes in structures, PC1, in each case. A representative structure from each set was analyzed by ANM to determine the global modes ANM1-ANM3. The ribbon diagrams (*left*) illustrate the global movements predicted by theory (*green arrows*) and exhibited by experiments (*violet arrows*). The plots on the *right* display the dispersion of the examined models/structures along these top-ranking mode axes derived from experiments (PC1) and theory (ANM 1, 2 or 3). A bound inhibitor colored gray is shown for HIV-RT, p38, and Cdk2 to highlight the binding pocket. The colored dots in the *right* plots refer to different types of structures, as labeled. Projected values, in Å, represent a collective distance/deviation for the entire structure. See Bakan and Bakar for details [125]. Results, diagrams and plots were generated using *ProDy* [155]. *Credit: the analysis and figure are attributed to Ahmet Bakan.*

3.4 ENM-BASED MODELING OF PROTEIN FLEXIBILITY AND STRUCTURAL CHANGES INVOLVED IN INHIBITOR BINDING

The previous section underscores the close overlap of the low-frequency modes of motions predicted by the ANM with the principal structural changes observed experimentally upon ligand binding. Here, we describe how theoretically predicted normal modes have been applied to modeling protein flexibility coupled to binding. Among these, two application patterns stand out: generating conformational ensembles for docking, and simultaneously docking the ligand and deforming the protein.

3.4.1 Generating conformational ensembles for docking

A practical approach to improving molecular docking is using an ensemble of conformations obtained by experimental or computational means [156]. Conformations can be generated using a number of web servers including ANM [157], ElNémo [158], FlexServ [159], Fiberdock [160], or using software packages like *ProDy* [155].While the number of modes to use for sampling or the relevance of the modes to the specific binding problem varies, the general recipe is to use a

small subset (up to 8-10) of soft modes, assuming that these will effectively map the most significant changes in structure. An alternative approach is to select the modes that are expected to facilitate ligand binding, e.g., those accompanied by local conformational changes at the binding site [161], or those accommodating the specific force applied by the ligand [162].

One of the earlier studies that took advantage of normal modes for generating alternative conformers to be then used in docking simulations is the work of Cavasotto, Kovacs and Abagyan [161]. The authors proposed a scoring mechanism for selection of ENM modes that are most relevant to the area of interest (e.g., binding site) by assessing the contribution of each mode to the deformability of that area. A scoring algorithm was applied to identifying a few modes, within the mid-frequency range, that best describe the loop flexibility in the binding pocket of cAMP-dependent protein kinase. To generate deformation vectors, these relevant modes were linearly combined by exhaustively sampling all combinations of uniformly discretized linear coefficients. After correcting unphysical geometries with energy minimization, the side-chains were optimized by docking known binders using a flexible side-chain docking algorithm. While this study demonstrated the utility of normal modes, it also raised concerns about the choice of normal modes, i.e., whether one needs to examine higher modes for accurate binding of ligands. The above calculations demonstrate however that inclusion of up to 20 modes usually provides a good description of structural changes associated with ligand binding. Further refinements (e.g., side chain isomerizations) would require energy minimizations or MD simulations with detailed potentials. In a more recent study from the same laboratory, Abagyan and coworkers [163] showed for a benchmark set of 28 bound structures that a consistent improvement in cross-docking results is achieved (compared to docking simulations with single experimental receptor conformation) when binding site ensembles generated with ENMs were

used. Likewise, Perahia and coworkers have made use of NMA to predict the binding modes of inhibitors in the active sites of matrix metalloproteinases (MMPs) correctly [164].

A more recent example is the work of Sperandio et al. [165] who explored the conformational space accessible to Cdk2 by deforming the structure along the first 25 lowest-frequency modes obtained from all-atom NMA (note that the lowest frequency modes are highly robust and reproducible with either full atomic or ENM-based models). For each mode, two deformations were generated in each direction, up to a mass-weighted RMSD of 2 Å. Thus a total of 50 deformed structures were generated. Unphysical geometries created by deforming the structure along each mode were eliminated by energy minimization. Since energy minimization may introduce artifacts, significantly altered conformations were discarded. The remaining conformations were then used to dock inhibitors of Cdk2, resulting in improved accuracy compared to the docking of the inhibitors onto an energy- minimized *apo* structure.

3.4.2 Coupled docking and deformation along normal modes

Current docking software handles a limited number of conformers, and hence limits the researcher to a small number of normal modes when generating ensembles. This has the potential shortcoming for highly flexible binding sites, as even for a small number of modes, let's say 6, the number of potential combination of these modes is $2^6 = 64$; and more importantly, the size of motion along a given mode usually varies depending on the ligand, as shown in recent examination of bound structures [125]. A computationally manageable approach would be to sample conformers at short intervals (e.g., total of 20 conformers, 10 in each direction, positive and negative), along each individual mode, but this approach has the potential problem of missing relevant conformations resulting from the combinations of modes. The alternate

approach is *in situ* utilization of normal modes, i.e., guiding protein energy minimization by normal modes during docking. As this approach can handle larger number of modes, it promises generating conformers specific to the ligand of interest.

Zacharias and Sklenar [166] proposed to use the deformation along the soft modes as variables in an energy minimization procedure that aimed at optimizing the steric complementarity of the receptor protein with a pre-docked ligand. The energy function consisted of an intermolecular interaction term and a receptor deformation penalty. Each mode contributed to the deformation penalty a quantity proportional to the fourth power of the magnitude of the deformation along that mode. The square of the eigenvalues was taken as force constants to account for the relative stiffness of the modes. The receptor deformation penalty thus calculated avoided the computationally more demanding calculation of the receptor intramolecular energy at each minimization step. The method was first applied to a DNA-ligand complex using the softest 40 modes calculated from all-atom NMA of the unbound DNA structure. The ligand was kept fixed in the known binding site, and the unbound DNA structure was deformed by energy minimization along the normal modes, resulting in a DNA structure with improved shape complementarity with respect to the ligand and more similar to the bound DNA structure. Subsequently, the method was applied to flexible protein-ligand docking using the soft modes derived from PCA of MD trajectories [167]. More recently, May and Zacharias used ENMs in place of all-atom models in NMA, and included the translational and rotational degrees of freedom of the ligand in the minimization scheme along the soft modes of the receptor structure [168,169]. Finally, side-chain flexibility was included at each minimization step with the help of a rotamer trial protocol [170]. This latter methodology has been tested [170] for two cases: docking known Cdk2 inhibitors to unbound kinase structures, and cross-docking of inhibitors to

several bound structures. Figure 11 illustrates the improvement in ligand placement achieved with the proposed methodology.

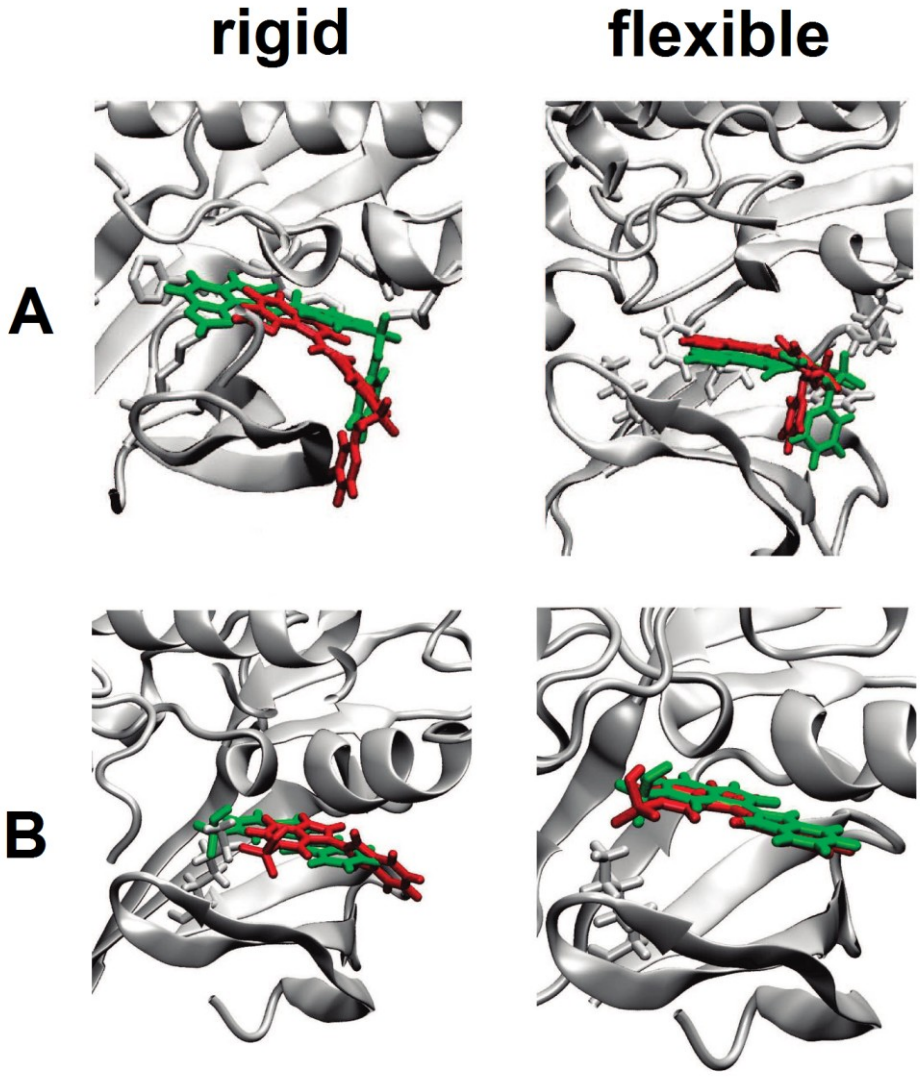

Figure 11. Improved docking of Cdk2 inhibitors by modeling the flexibility of Cdk2. Cdk2 (gray cartoon), experimental inhibitor binding mode (green), docked inhibitor binding mode (red). (A) Ligand from PDB 1E9H docked to apo structure of Cdk2 (PDB 1HCL) using rigid receptor (left) and flexible receptor modeling (right). (B) Cross-docking of inhibitor from PDB 1FVV to another inhibitor bound structure of Cdk2 (PDB 1E1V) using rigid receptor (left) and flexible receptor modeling (right). Adapted with permission from May and Zacharias [170] - Copyright 2008 American Chemical Society.

Lindahl and Delarue [171] have proposed an alternative strategy for refining protein-ligand and protein-DNA structures: to optimize the deformation along each mode, one at a time, by scanning the mode amplitudes and calculating the resulting intermolecular energy. They have found that pre-sorting the modes in order of single-mode largest energy reduction achieved better results than using them in order of mode frequency (see Figure 12). The modes in this study were calculated using all-atom ENM, thus the positions of side-chain atoms were directly obtained, with the caveat that the model would not account for rotameric transitions in side-chains.

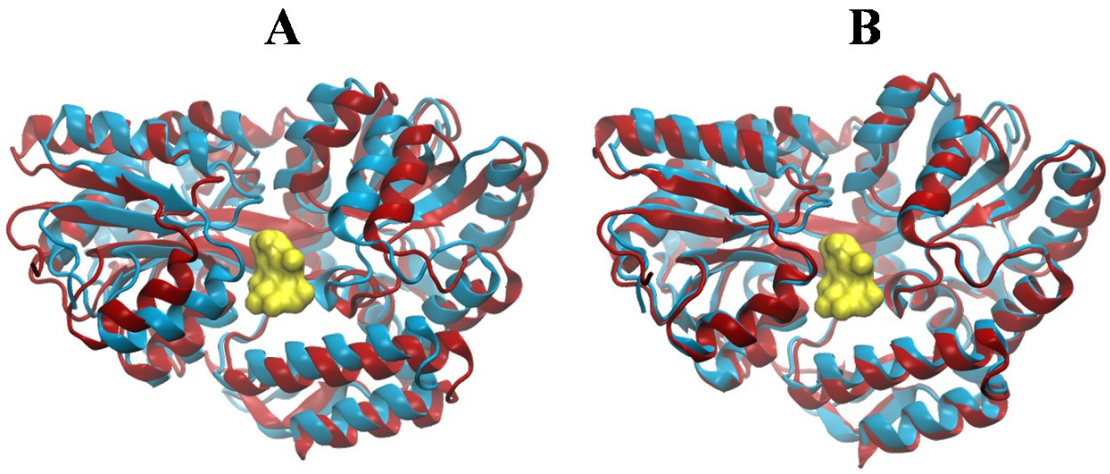

Figure 12. Refinement of maltodextrin binding protein along low-frequency normal modes. (A) The unbound structure (red; PDB 1OMP) is superimposed onto the ligand (yellow) bound structure (blue, PDB 1ANF). The initial RMSD is 3.77 Å (B) After refinement of the unbound form along the top 5 lowest frequency (all-atom) ENM modes, Lindahl and Delarue [171] obtained a structure with an RMSD of 1.86 Å from the experimentally known bound form. Shown is a similar structure (RMSD of 1.51Å) obtained by projecting the deformation vector along the two lowest-frequency ANM modes.

Mashiach, Nussinov and Wolfson [160,162] have proposed a docking refinement protocol based on a new strategy for selecting the most relevant normal modes; namely, those modes that exhibit the highest correlation with the direction of the repulsive van der Waals (vdW) forces the ligand exerts on the protein. Energy minimization along these modes relaxes the protein to accommodate the ligand. This strategy allows for iterative selection of relevant

normal modes from an unlimited number of *a priori* modes. The application to 20 protein-protein complexes demonstrated the utility of the approach for improving the accuracy of docked conformers and obtaining the correct ranking of near-native docking solutions [162].

3.5 FINAL REMARKS: PRE-EXISTING SOFT MODES OF MOTION FACILITATE LIGAND BINDING TO PROTEINS

An alternative view in addition to the three shown in Figure 9 is suggested by the wealth of computational data presented above to be in accord with experimental data. This view, shortly termed *pre-existing paths*, as opposed to pre-existing substates, is based on the shape/curvature of the energy landscape near the native state of the protein. It considers the fluctuations near the global energy minimum, rather than the jumps between the local minima. The energy surface has different curvatures along each of the $3N$-6 collective degrees of freedom (or collective fluctuation directions) that define various reconfiguration mechanisms (for a structure of N sites/nodes). The intrinsic motions thermally accessible to the protein are simply fluctuations along these collective directions, or modes, each of which defines an uphill path of reconfiguration away from the native state (bottom of the energy well). A multitude of uphill paths are therefore accessible during the collective fluctuations due to the stochasticity of thermal motions, those with softer curvature being more probable and larger in amplitude. Numerous applications including those summarized here have suggested that these very fluctuations along the soft modes are (i) uniquely encoded by the native structure, (ii) robustly computed using simple models, such as the ANM exclusively based on native contact topology and (iii) functionally relevant, e.g., they enable the opening and closing of the binding site, they

trigger allosteric switch mechanisms, and even facilitate the stabilization of catalytically competent/pre-disposed states. A ligand in the vicinity of a protein undergoing these fluctuations would selectively recognize an instantaneous conformer that provides an optimal binding site, and stabilize this transient conformer into a new bound substate. The encounter with the ligand might thus change the energy landscape (originally an uphill path) into a stable intermediate state (local energy minimum) or equilibrium state (a new global energy minimum), even if that particular conformer was *not* a pre-existing 'substate', but simply accessible via a pre-existing 'soft path'.

4.0 LIGAND-BASED DRUG DESIGN: A NEW CLIQUE-BASED APPROACH TO BIOACTIVE CONFORMATION IDENTIFICATION

In this chapter we present a new approach for generating models of the bioactive conformation of molecules in the absence of protein structure, the validation and selection of the 'best model' in accord with experimental data, and its use to identify new bioactive compounds. We also show the application of the method to known c-Myc/Max inhibitors.

Pharmacophore-based superposition of bioactive compounds is a popular approach to generate plausible models of their bioactive conformations. The underlying assumption is that bioactive molecules share a common binding mode to a - possibly unknown - biological target, or, in other words, bioactive molecules share a common spatial distribution of compatible chemical/pharmacophore features. Pharmacophore-based methods attempt to superimpose the bioactive molecules such as to maximize the overlay of common 3D pharmacophore features. However, a problem with that approach occurs when the molecules are similar; in this case, multiple superpositions are possible, raising doubts about the 'correct' superposition model to use. It is therefore important to generate a list of candidate superposition models and to select the one that correlates best with experimental data.

4.1 PROBLEM FORMULATION

The input to the algorithm is a set of M bioactive molecules, with at most C pre-generated conformations per molecule. The output is a list of N superposition models. These models are then tested for their ability to reproduce known experimental data, such as the ability to discriminate between known active and inactive compounds. Finally, the best model is selected and used to screen for new active compounds in molecular databases.

Typically, putative bioactive conformations are identified after optimal superposition of the molecules. Our strategy works in the opposite direction: first we identify a set of (putative) bioactive conformations, one conformation per molecule, to subsequently generate a superposition model. Note that since there may be up to C^M possible combinations of conformations, it would be impractical to exhaustively evaluate all possible combinations. For example, in a typical scenario involving 5 molecules, each with 100 conformations, there would be 10^{10} possible ways to select a set of conformations. Most importantly, what would be the criterion to label a set of conformations as putative bioactive?

The answer to the above question paves a way for solving the problem efficiently. Under the common binding mode assumption, the bioactive conformations should be mutually similar, in terms of both their shape and chemical features. Thus, by finding mutually similar conformations, we should, in principle, identify putative bioactive conformations. This insight reduces the problem to selecting one conformation for each molecule such as to maximize the sum of pairwise similarity scores among the selected conformations. In turn, this can be solved efficiently using a maximum-edge weight clique algorithm.

Let G be a graph in which the vertices represent the individual conformations. An edge connects two vertices if, and only if, the corresponding conformations belong to different

molecules. The weight associated with each edge is the shape- and chemistry-based similarity score between the two conformations. Finding a set of conformations that maximize the sum of pairwise similarity scores corresponds to finding a clique of size *M* with the maximum sum of edge weights (see Figure 13). The next section describes the algorithm for finding such a clique.

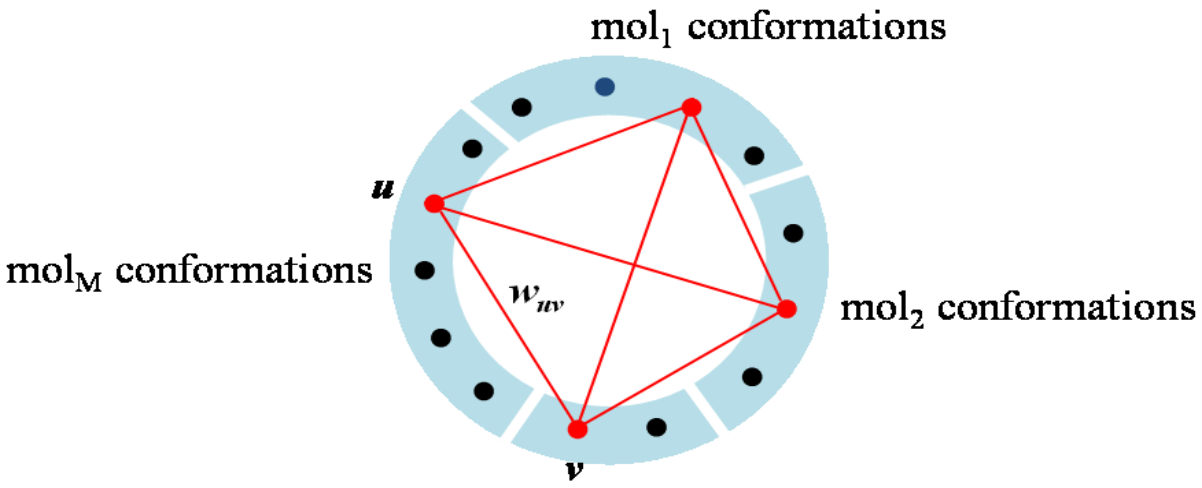

Figure 13. Reduction of the bioactive conformation identification problem to the maximum edge weight clique problem. The vertices in the graph represent conformations. There is an edge connecting two conformations, if they are from different molecules (not all edges are shown for simplicity). Finding a set of conformations that maximize the sum of pairwise similarity scores corresponds to finding a clique (red color) of size *M* with the maximum sum of edge weights (similarity scores).

4.2 ALGORITHM

The maximum edge weight clique problem is a well-known NP-hard problem in computer science. Here we describe an integer linear programming (ILP) formulation of the problem which was inspired by an ILP formulation of the protein side-chain packing (SCP) problem in computational biology [172]. The SCP problem is related to our formulation of the bioactive conformation identification problem, as they both are combinatorial optimization problems with

objective functions expressed as a sum of pairwise scores (side-chains interaction energy in the SCP problem). The realization of this connection is important because techniques applied to one problem are transferable to the other. For instance, the Dead End Elimination [173] algorithm, which has been successfully applied to the SCP problem, could also be useful to the bioactive conformation identification problem. Our decision to use the ILP formulation was motivated by the ease of implementation and availability of standard ILP solvers.

In the ILP formulation, we associate a binary decision variable {0,1} to each vertex and edge of the graph. Let x_u be the binary variable associated to vertex u, and x_{uv} be the binary variable associated to edge (u,v). If vertex/conformation u is selected, then x_u=1; otherwise, x_u=0. If both vertices u and v are selected (x_u = 1 and x_v = 1), then x_{uv} = 1; otherwise, x_{uv} = 0. With this coding scheme, the objective function can be written as max $\sum_{(u,v)} w_{uv}x_{uv}$ where w_{uv} is the weight of edge (u,v). To ensure that only one conformation per molecule is selected, we add one constraint for each molecule j=1,…,M: $\sum_{u\in mol_j} x_u = 1$. In addition, to ensure the correct assignment of edge variables induced by assignments of vertex variables, we add the following additional constraints [172]:

Equation 5

$$\sum_{u\in mol_j} x_{uv} = x_v \quad j = 1, \dots, M \text{ and } v \in \bigcup_{k\neq j} mol_k$$

where $v \in \bigcup_{k\neq j} mol_k$ denotes that v is selected from all molecules other than j. This constraint demands that if a vertex is not selected (x_v = 0), then no adjacent edge can be chosen; if a vertex is selected (x_v = 1), then exactly 1 adjacent edge is chosen for each vertex/conformation set. This ensures the selection of edges that are in the subgraph induced by the choice of conformations. The ILP formulation is shown below:

Equation 6

$$\max \sum_{(u,v)} w_{uv} x_{uv}$$

subjected to

$$\sum_{u \in mol_j} x_u = 1 \qquad j = 1, \ldots, N$$

$$\sum_{u \in mol_j} x_{uv} = x_v \qquad j = 1, \ldots, N \text{ and } v \in \bigcup_{k \neq j} mol_k$$

$$x_u \, , \, x_{uv} \in \{0,1\}$$

With this formulation, putative bioactive conformations can be found using any ILP solver software, such as Gurobi [174], CPLEX [175] and GLPK [176]. In the present study we used the software Gurobi 4.0. To find several optimal or near optimal solutions, the ILP problem is solved multiple times, at each new iteration the last solution found is excluded by adding the constraint [172]: $\sum_{u \in S} x_u \leq M - 1$, where $S = \{ u \mid x_u = 1 \}$ is set of conformations selected in the previous iteration. This requires that at the next iteration a new solution differs from previous solutions by at least one conformation.

4.2.1 Calculation of the similarity score between two conformations

Prior to solving the ILP problem, we need to compute the weights w_{uv} of all edges in the graph, that is, the similarity score between all pairs of conformations from different bioactive molecules. To this end, we use the program ROCS 3.1.2 (OpenEye Scientific Software) for shape-based superposition of molecules [177-179]. In ROCS, the shape of a molecule corresponds to the volume occupied by its atoms in 3D. To superimpose two molecules based on their shapes, ROCS maximizes the intersection volume (overlap) of the two molecules.

The most striking feature of ROCS is how it efficiently computes, analytically, volume overlaps and their gradients with respect to the (x,y,z) coordinates. This is achieved by using Gaussian functions centered at the atoms, in lieu of the hard spheres commonly used to represent the vdW volume of atoms [177-179]. The smoothness of the gradients introduced by Gaussian functions avoids numerical instabilities in the optimization of the rigid-body transformation that maximizes the overlap of the two molecules. Next we briefly review the theory behind the shape-based superposition of ROCS using Gaussian functions.

The volume of a hard sphere of radius *R* is $\frac{4}{3}\pi R^3$, which is theoretically calculated from the integral:

Equation 7

$$V_{sphere} = \int f(r_i)\boldsymbol{dr} = \frac{4}{3}\pi R^3$$

$$f(r_i) = \begin{cases} 1 & r_i \leq R \\ 0 & r_i > R \end{cases}$$

where the integral is over the whole space, r_i=$|\boldsymbol{r}-\boldsymbol{r_i}|$ is the distance from the sphere center $\boldsymbol{r_i}$, and $f(r_i)$ is the *volume characteristic function.* The intersection volume of two hard spheres can be computed from their respective characteristic functions, f_1 and f_2:

Equation 8

$$V_1 \cap V_2 = \int f_1(r_1) f_2(r_2)\boldsymbol{dr}$$

It is clear that the product $f_1(r_1)f_2(r_2)$ equals 1 only where the two spheres overlap and 0 otherwise; therefore, Equation 8 corresponds to the volume of the overlap of two spheres by analogy to Equation 7. Grant and Pickup [177] replaced the step function *f(r)*, characteristics of a

hard sphere, by a Gaussian function with two defining parameters: its prefactor p and its width w (see Equation 9). They found that the volume of a hard sphere could be recovered to high accuracy (~ 0.1%) by fixing p to 2.7 and setting w for each atom such that the volume integral agreed with its hard-sphere volume.

Equation 9

$$\rho(r) = pe^{-wr^2}$$

Since the product of two Gaussians is another Gaussian, the intersection volume calculated from Equation 8 has a simple analytical solution. Grant and Pickup [177,178] provided analytical equations to compute the intersection volume of N Gaussians ($V_1 \cap V_2 \cap ... V_N$) as well as and the gradient of the intersection volume with respect to the coordinates. These equations are written as functions of the parameters of each Gaussian (p, w and the coordinate center). The calculation of intersection volumes, $V_1 \cap V_2 \cap ... V_N$, are used to compute the intersection volume of two molecules. First note that the volume of a molecule with two atoms is given by

Equation 10

$$V_1 \cup V_2 = V_1 + V_2 - V_1 \cap V_2$$

and the volume of a molecule with N atoms, i.e., a generalization of the above, is given by:

Equation 11

$$\bigcup_{i=1...N} V_i = \sum_i V_i - \sum_{i<j} V_i \cap V_j + \sum_{i<j<k} V_i \cap V_j \cap V_k - \sum_{i<j<k<l} V_i \cap V_j \cap V_k \cap V_l + \cdots$$

Note that we need to take into account double, triple, and higher order intersections between atoms. Given two molecules A and B, their volumes are computed as in Equation 11 and the intersection volume as in Equation 12:

Equation 12

$$V_A \cap V_B = \sum_{i \in A, j \in B} V_i \cap V_j - \sum_{i,j \in A, k \in B} V_i \cap V_j \cap V_k - \sum_{i,j \in B, k \in A} V_i \cap V_j \cap V_k + \cdots$$

where all the intersection terms up to the sixth order and between atoms within a cutoff distance are computed analytically.

Four starting points for optimization are generated by aligning the molecules' centroids and their two largest principal components (each pair of principal components can be aligned in two directions). ROCS then performs a gradient-based maximization of the volume overlap using a quaternion parameterization of the rotation angles [178]. The superposition with highest volume overlap is reported out of the four maximization runs. In addition to shape, ROCS also attempts to maximize the overlap of 'compatible' chemical features, which are also modeled as Gaussian functions. Among the several scoring functions reported by ROCS, we use the TanimotoCombo score, because it combines both the shape and chemistry superposition of the conformations into a single score.

4.2.2 Generation of superposition models

A solution of the ILP problem is a set of *M* conformations selected for being most mutually similar. Next we describe how these conformations are all superimposed. We superimpose *M* conformations by performing *M*-1 pairwise superpositions using ROCS [178] (see Section 4.2.1). One simple way to accomplish this is to designate one 'reference' conformation and then

superimpose the remaining *M*-1 conformations to it. However, it is more desirable to find the set of *M*-1 superpositions that maximize the sum of superposition scores. These *M-1* pairwise superpositions correspond to the edges of the maximum spanning tree of the clique induced by the *M* conformations (Figure 14). For each edge of the maximum spanning tree, a superposition of the adjacent conformations is performed using ROCS [174]. Note that the order by which the superpositions are carried out is irrelevant. Note also the following transitive effect: if you superimpose A to B and B to C, then you automatically have A superimposed to C.

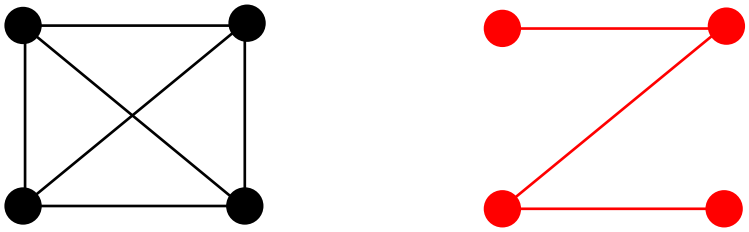

Figure 14. (Left) Clique formed by 4 conformations. (Right) The maximum spanning tree of the clique defines the best 3 pairwise superpositions to overlay 4 conformations. See the text for details.

4.3 APPLICATION TO C-MYC/MAX DISRUPTORS

Here we report the application of the proposed methodology to modeling the bioactive conformations of known c-Myc/Max inhibitors, thus helping to delineate the shape and chemical features responsible for biological activity. There is no experimental structure of c-Myc in complex with small molecules. Our experimental dataset consists of 23 related compounds, of which 14 are actives and 9 inactives. The five compounds shown in Figure 15 were selected among the most active compounds to generate the model. We pre-generated at most 200 conformations per compound using the software Omega2 2.4.3 (OpenEye Scientific Software)

[180]. Our program analyzed these conformations and outputted 30 superposition models (user-specified parameter).

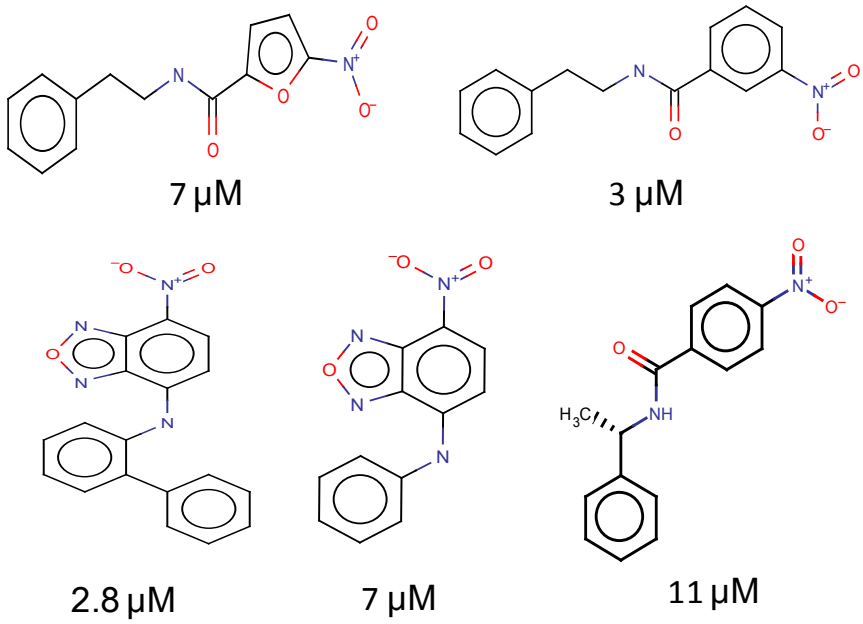

Figure 15. Selected c-Myc/Max inhibitors and their *in vitro* binding affinity (K_d).

We select the best model based on its ability to discriminate between active and inactive compounds. The area under the ROC (Receiver Operating Characteristic) curve provides a metric for model comparison. We use each superposition model as a query in the program ROCS. ROCS ranks order the 23 compounds in our dataset according to their overlay score (TanimotoCombo) to the model and then it computes the area under the ROC curve (AUC). The AUC ranged from 0.751 (model 27) to 0.896 (model 22) – see Figure 17. The average AUC over the 30 models was 0.799. The best model (model 22) is shown in Figure 16.

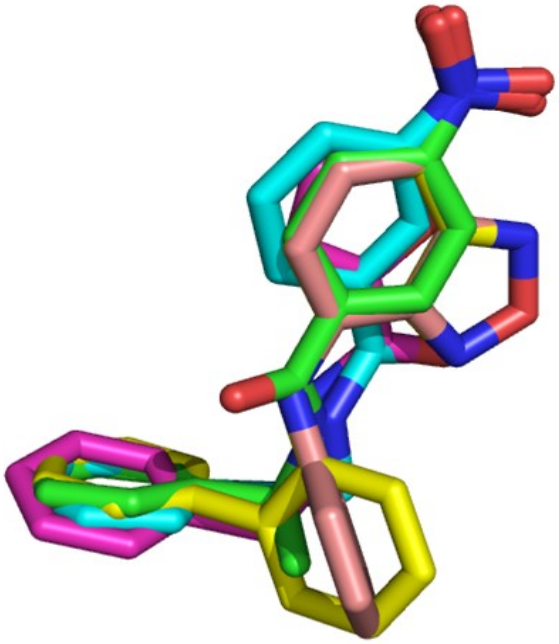

Figure 16. Superposition model 22.

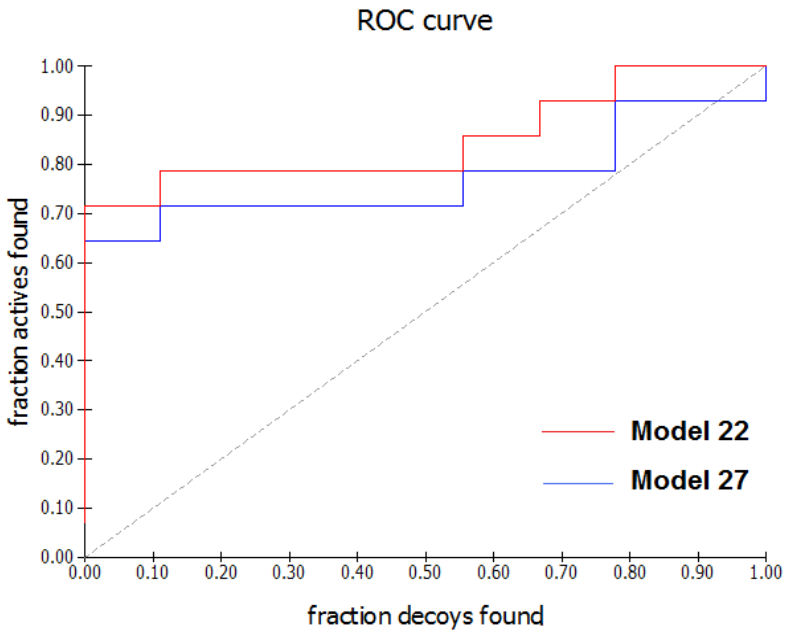

Figure 17. ROC curve of model 22 (the best) and model 27 (the worst) are 0.896 and 0.751, respectively.

Next we calculated the AUC obtained by using each one of the 5 conformations of model 22, individually, as a query. The AUC ranged from 0.576 (Mol 1) to 0.777 (Mol 4) – see Figure 18. Interestingly, using the entire ensemble of conformations as a query outperformed the results obtained by using each conformation individually. A possible explanation for that involves the density of atoms superimposed in the ensemble. Areas with high density of atoms weight more, and, therefore, molecules that do not overlay well in those areas are penalized. For instance, note in Figure 16 that the nitro group present in all 5 molecules overlay almost perfectly, hence this model penalizes a molecule if its nitro, or similar, group deviates from this 'optimal' position. Likewise, the phenyl groups superimposed in the leftmost part of Figure 16 are all in the same plane, thus the model penalizes a molecule with phenyl, or similar, group lying out of that plane.

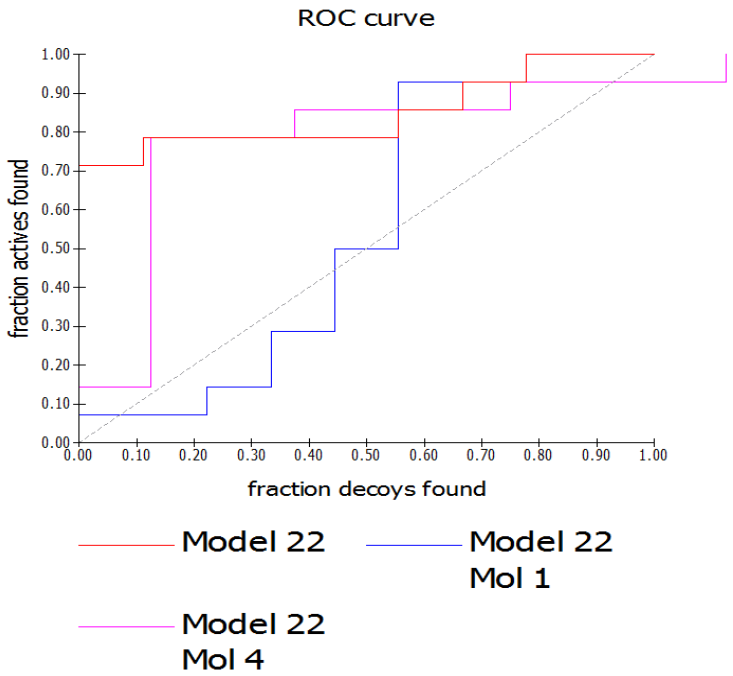

Figure 18. Comparison of results obtained using the entire ensemble vs. individual conformations as queries.

Experimental Validation

We used Model 22 (Figure 16) as ROCS query against the molecular libraries of the U. of Pittsburgh Center for Chemical Methodologies and Library Development (PITT CMLD) [181]. We selected 29 compounds, among the top hits, for experimental testing at Prof. Prochownik's Lab. At the time of this writing, the experiments were ongoing but we obtained confirmatory activity of at least one compound. Figure 19 shows the results from an Electrophoretic Mobility Shift Assay (EMSA) which quantitatively measured the dissociation of DNA from c-Myc/Max complex for different compound concentrations.

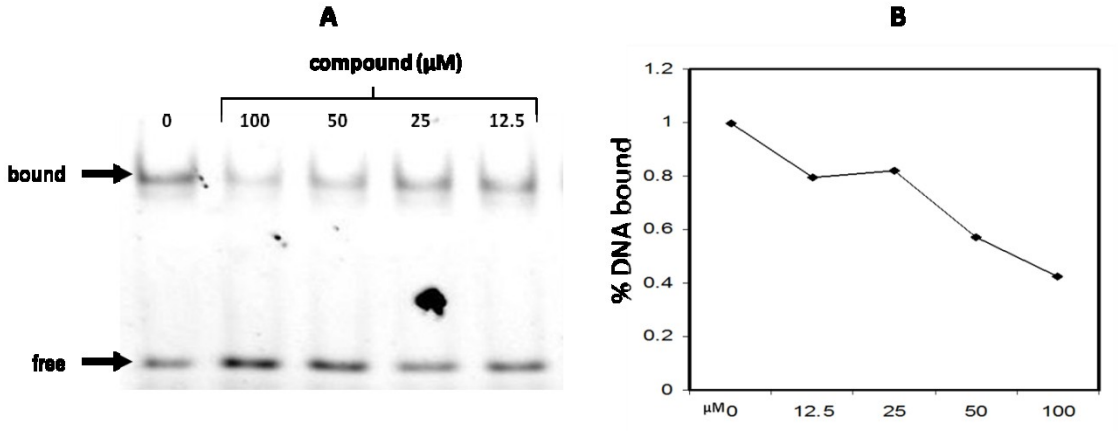

Figure 19. (A) EMSA results. Recombinant c-Myc353-439 and full-length Max(S) were purified to homogeneity from E. coli and used at a final concentration of 30 nmol/L in the presence of the indicated concentration of compound. A 22 bp E-box-containing dsDNA oligonucleotide labeled on one strand with FAM was used at 10 nmol/L concentration in all reactions. B. Quantitative analysis.

5.0 STRUCTURE-BASED DRUG DESIGN: MIMICKING PROTEIN-PROTEIN INTERFACES WITH SMALL DESIGNED MOLECULES

One way of designing small-molecule inhibitors of PPIs is to replicate the binding epitopes of one of the protein partners onto a small-molecule scaffold. In this chapter, we describe how the structure of the c-Myc/Max complex (PDB: 1NKP) was exploited to design compounds that mimic the localized binding interactions that Max makes with c-Myc. To achieve this goal, we developed a semi-automated protocol as described below.

One potential problem with Max mimetic compounds is that since c-Myc is intrinsically disordered prior to complexation with Max, a small-molecule Max mimetic may not be able to overcome the high entropic barrier associated with inducing an ordered state on c-Myc. In anticipation of this potential problem, we adopted the following strategy: (i) we favor semi-rigid molecular designs with high degree of pre-organization of the binding groups, so as to reduce the entropic loss occurring upon binding; (ii) we target the hydrophobic regions of c-Myc since those tend to be less disordered; (iii) we choose a binding site confined to a short contiguous segment along the amino acid sequence of c-Myc. The latter implies that only a small segment of c-Myc will be required to fold upon binding, thus reducing the entropic cost that would otherwise be incurred if a larger portion of the protein was required to fold. This is supported by Hammoudeh *el al.* [28] findings that binding of c-Myc inhibitors at three independent sites induced only local conformational changes and preserved the overall disorder of c-Myc.

5.1 SELECTION OF BINDING SITE IN C-MYC

We used the software FRED Receptor 2.2.5 (OpenEye Scientific Software) to identify candidate binding sites on c-Myc. The methodology implemented in FRED Receptor 2.2.5 allows for docking of small-molecule probes around the protein surface and then identifying binding sites as highly populated clusters of probes. The predicted binding sites are shown in Figure 20. We evaluated these sites using the criteria outlined in the previous section (Section 5.0). Note that site 2 requires the folding/packing of the helix-loop-helix segments, thus targeting this site is likely to be entropically expensive (unless the packing interactions are pre-existing, or accessible via soft modes, in free c-Myc, but there is no experimental evidence of that). Likewise, site 3 appears between a helix and loop, and may also be entropically expensive. Besides that, sites 2 and 3 are located away from the interface of the c-Myc/Max complex; targeting these sites may either disrupt c-Myc/Max through an allosteric effect or it may actually stabilize the complex. For instance, Jiang *et al.* [30] discovered stabilizers of the Max homodimer by targeting a site in Max corresponding to site 2 in c-Myc - presumably, the compounds stabilized the helix-loop-helix packing in the bound form of the Max/Max complex. The only site remaining, i.e. site 1, appears to be a good site to target, since it is localized, hydrophobic, and situated in the c-Myc/Max interface between the leucine zipper domains of c-Myc and Max. Therefore, we selected site 1 and we designed compounds to mimic the interactions of Max with c-Myc at this particular site.

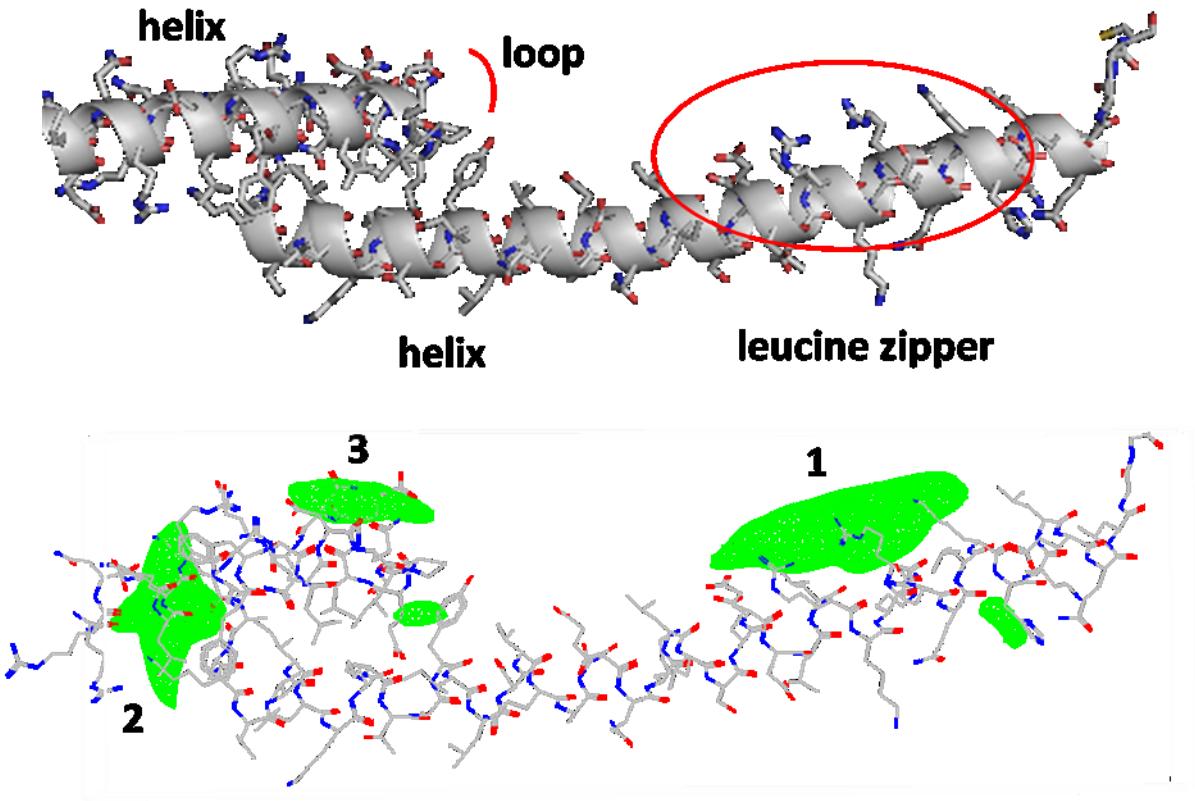

Figure 20. Predicted small molecule binding sites (green mesh). Binding site 1, located in the leucine zipper domain, was selected. See text for details.

5.2 MOLECULAR MIMICRY DESIGN PROTOCOL

We developed a semi-automated protocol to facilitate the design of Max mimetic compounds. In this protocol, we specify what we want to mimic by pre-positioning molecular fragments in 3D space and/or providing desired pharmacophore features. Molecular mimicry is achieved by incorporating the fragments into the designed molecules while keeping their pre-determined positions, and also by satisfying the pharmacophore features provided. Fragments and pharmacophores, representing precisely the features to be mimicked, should be carefully selected to map the essential chemical features required for binding. For mimicking a protein-protein

interface, fragments and pharmacophores can be taken from hotspot residues, which can also be used to guide the selection of docked fragments.

The goal was to generate molecules that incorporate the selected fragments and pharmacophores using small-molecule scaffolds (e.g. combinatorial chemistry scaffolds). However, scaffolds will seldom possess the exact geometry required to simultaneously connect multiple fragments and satisfy the pharmacophores. To overcome this limitation, we divided the process into two steps (see Figure 21). The first step was to anchor the scaffold on the binding site by any suitable means. We accomplished that by using one or more of the following methods: (1) docking the scaffold on the binding site; (2) directly connecting the scaffold to one pre-positioned fragment (shown in Figure 21, Step 1); (3) manually positioning the scaffold.

Once the scaffold was anchored on the binding site, the next step was to build the scaffold side chains. We used fragment-based drug design techniques, fragment linking and fragment growing, to link a side chain to a nearby fragment, or to grow a fragment from a side chain to meet a nearby pharmacophore feature (see Figure 21, Step 2). Fragment linking and growing were accomplished using the software ReCore (BioSolveIT) [182], which returns, in a matter of seconds, fragments satisfying one or more bond vectors and pharmacophores. Note that scaffolds and fragments are annotated with attachment points (dummy atoms) to indicate the positions available for chemical linkage. The resulting molecules are derivatives of the employed molecular scaffolds, and the side chains of which are formed by the union of linkers to pre-positioned fragments (fragment linking) or extensions that satisfy the pharmacophores (fragment growing).

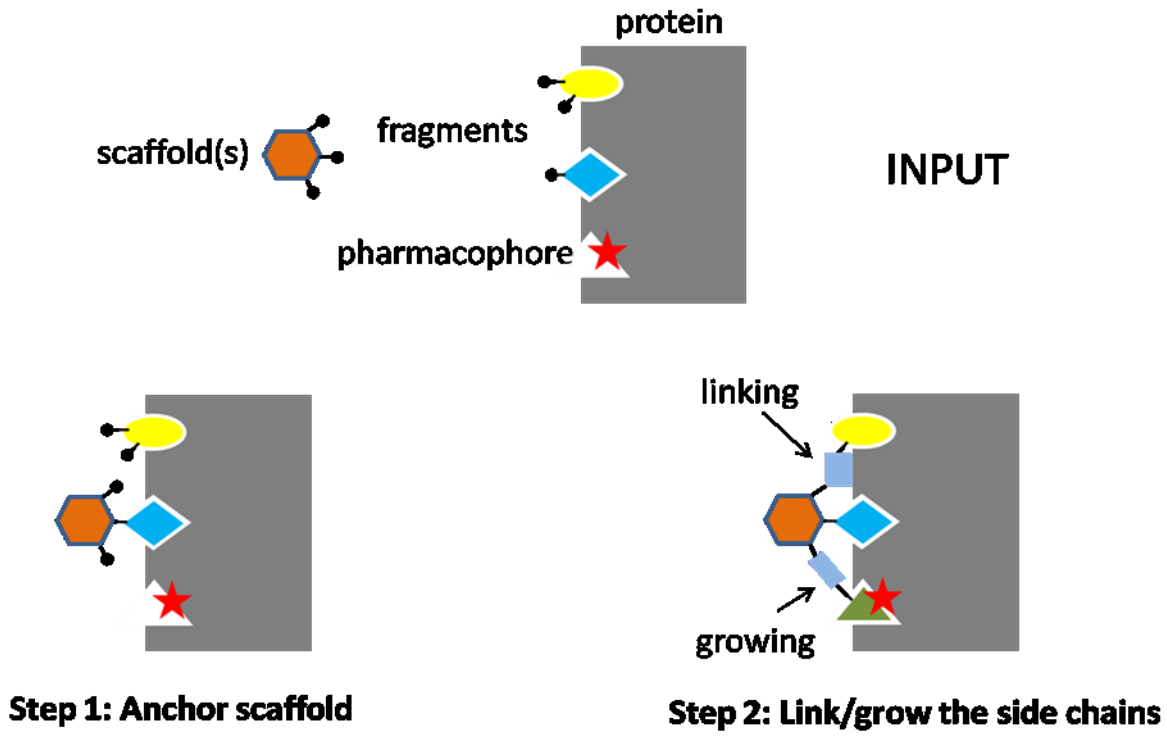

Figure 21. Sketch of the protocol developed for mimicking protein-protein interfaces with small designed molecules.

The algorithmic implementation of the proposed protocol is very challenging. We found a compromise between the convenience of automation and the human intervention required to design meaningful, synthetically feasible, compounds. While human intervention allows the user to incorporate prior knowledge and insights, which are currently impossible to automate, it also introduces subjectivity and, therefore, the results obtained may not be easily reproducible. The table below shows the division of steps between the computer and the user.

Table 1. Semi-automated implementation of the molecular mimicry design protocol. Division of steps between the computer and the user.

User Intervention	Computer Automation
1. Decision of what to mimic, and selection of pre-positioned fragments and pharmacophore annotations accordingly;	3. Scaffold anchoring/placement by docking or connecting the scaffold directly to pre-positioned fragments; filtering and ranking of scaffold placements by interaction energy;
2. Selection of small-molecule scaffolds;	5. Link/grow scaffold side chains to nearby fragments/pharmacophores and generate candidate molecules; ranking the candidate molecules by interaction energy;
4. Selection of scaffolds' placements from the previous step and/or manual scaffold placement;	7. Energy minimization of the molecules in the binding site. Re-start from step 1 if no satisfactory result was obtained.
6. Selection of candidate molecules; optionally, replace a fragment using the software ReCore [182].	

5.3 COMPOUNDS GENERATED USING THE PROPOSED PROTOCOL

Figure 22 shows four examples of molecules designed to mimic the interactions that Max makes on the selected binding site in the leucine zipper domain of c-Myc. Note the shape complementarity of the molecules to the binding site. Besides, the molecules make similar contacts to c-Myc as those made by Max residues. In total, 18 molecules were designed using molecular scaffolds from the University of Pittsburgh Center for Chemical Methodologies and Library Development (CMLD) [181]. Compounds similar to the ones generated using the proposed protocol were identified through 3D similarity searches (OpenEye ROCS software) from the CMLD library and sent for experimental testing.

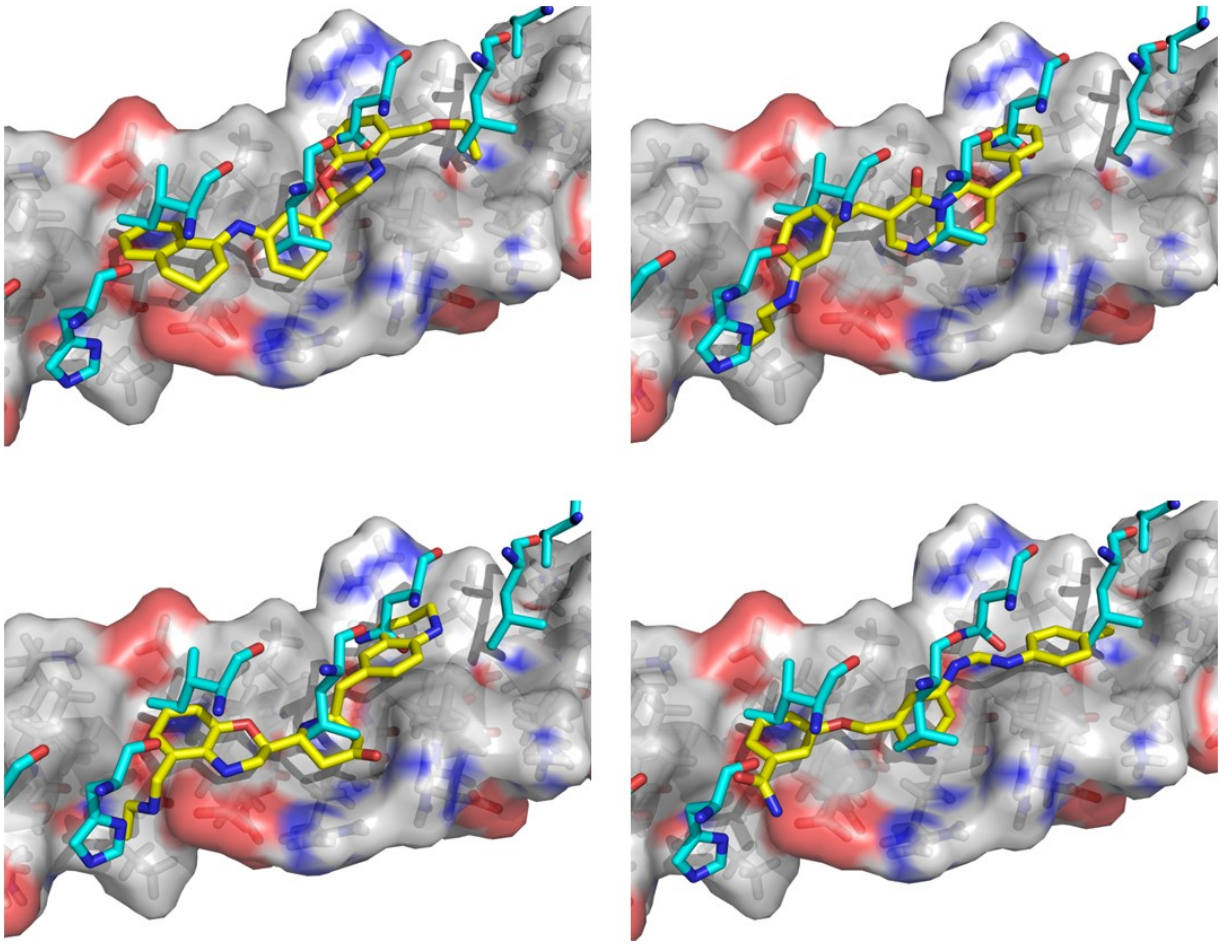

Figure 22. Molecules (yellow) designed to mimic the interactions that Max (cyan residues) makes with c-Myc (surface).

6.0 CONCLUSION

6.1 CONTRIBUTIONS MADE TOWARDS FACILITATING THE DISCOVERY OF INHIBITORS OF PROTEIN-PROTEIN INTERACTIONS

The discovery of small-molecule inhibitors of PPIs is a challenging endeavor pursued by industry and academic institutions. In this PhD dissertation, we aimed at assisting in increased productivity in the field through the following contributions:

We presented a thorough review of current paradigms and strategies for discovering small-molecule modulators of PPIs (Chapter 2.0). We briefly introduced each approach, critically analyzed their advantages and disadvantages, and provided examples, taken from recent literature, of their successful applications. We also explained the therapeutic motivation for targeting the PPIs discussed in the examples. Since there is no single best drug discovery paradigm capable of tackling all drug targets, it is important to understand the strengths and weaknesses of each approach.

In Chapter 3.0 , we presented an extensive analysis of the role of protein flexibility in molecular recognition and binding, and reviewed recent applications of ENM-based normal mode analyses to modeling the conformational changes of protein observed upon inhibitor binding. We classified the applications in two groups, namely: (1) generation of protein conformational ensembles for docking, and (2) coupled docking and deformation along normal modes. In addition, we proposed that *pre-existing soft modes* of motion, thermally accessible by

the protein, facilitate ligand binding. Chapter 3.0 was based on an article that has been recently accepted for publication in *Protein Science* [124].

We developed a new algorithm for generating putative models of the bioactive conformations of molecules in the absence of protein structure (Chapter 4.0). We showed how to validate and select the 'best model', according to its ability to discriminate between known actives and inactives. Our method exploits a new insight: we can identify bioactive conformations by identifying the set of conformations that contains mutually similar members, in terms of their shape and the spatial distribution of their chemical features. To solve the problem we implemented an Integer Linear Programming formulation of the maximum edge weight clique problem. Using a model of the bioactive conformations of known c-Myc/Max inhibitors generated by this method, we have identified a new c-Myc/Max inhibitor confirmed experimentally and several other potential inhibitors to be tested experimentally.

In Chapter 5.0 , we showed how the structure of a protein-protein complex can be exploited to design compounds that mimic the binding epitopes of one of the proteins. We designed compounds that mimic localized interactions of Max at the leucine zipper domain of c-Myc, and identified similar compounds from the CMLD library that will be tested experimentally.

6.2 FUTURE DIRECTION

We envision some possible directions for future extensions and enhancements of the work herein presented. We believe the potential of ENMs for modeling protein flexibility coupled to ligand binding is an area that could be further exploited. The work reviewed in Chapter 3 provides

encouraging results, but it still lacks integration of ENMs with existing molecular docking software. As a step towards making ENMs more accessible to the general scientific community, we have developed *ProDy* [155], a Python-based software package that allows non-expert users as well as advanced software developers to take advantage of ENMs functionality with minimum effort [Credit: *ProDy*'s primary developer was Dr. Ahmet Bakan].

One possible extension of *ProDy* concerns the implementation of more advanced algorithms for sampling protein conformations along relevant normal modes. The deformation of protein structures along normal modes generates unphysical geometries, which are usually corrected by energy minimization. However, energy minimization is computationally expensive and often causes significant distortions to the protein structure. An improved approach would consist of applying normal mode deformations only to C_α atoms and then making use of all-atom protein reconstruction algorithms [183-185]. These algorithms are capable of generating a full atomistic model of the protein structure from its reduced or coarse-grained representation, such as approximate position of backbone and/or side chains atoms. This approach promises to efficiently generate protein conformations guided by normal modes and free of unphysical geometries.

In Chapter 4, we used an experimental dataset consisting of both active and inactive compounds to validate and select the best superposition model of the bioactive conformations of related c-Myc/Max inhibitors. However, in many practical circumstances, inactive compounds that share similar chemotypes may not be available to compute the area under the ROC curve utilized for model selection. To consider this scenario, the method could be extended to search for *robust* maximum edge weight cliques, which are insensitive to small changes in the weights.

The molecular mimicry design protocol presented in Chapter 5 could benefit from further automation, although we still expect some degree of human intervention be required to design meaningful, synthetically feasible, compounds. An integrated graphical user interface could, nevertheless, greatly facilitate the design process. As currently implemented, the attachment of fragments to a molecular scaffold, either by linking or growing, is unrestrained, but it would be useful to incorporate chemical rules to reflect the constraints imposed by synthetic organic chemistry.

BIBLIOGRAPHY

1. White AW, Westwell AD, Brahemi G (2008) Protein-protein interactions as targets for small-molecule therapeutics in cancer. *Expert Reviews in Molecular Medicine* 10: 1-14.

2. Fry DC, Vassilev LT (2005) Targeting protein-protein interactions for cancer therapy. *Journal of Molecular Medicine* 83: 955-963.

3. Arkin M (2005) Protein-protein interactions and cancer: small molecules going in for the kill. *Current Opinion in Chemical Biology* 9: 317-324.

4. Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 450: 1001-1009.

5. Arkin MR, Wells JA (2004) Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nature Reviews Drug Discovery* 3: 301-317.

6. Hopkins AL, Groom CR (2002) The druggable genome. *Nature Reviews Drug Discovery* 1: 727-730.

7. Berg T (2003) Modulation of protein-protein interactions with small organic molecules. *Angewandte Chemie International Edition in English* 42: 2462-2481.

8. Fry DC (2006) Protein-protein interactions as targets for small molecule drug discovery. *Biopolymers* 84: 535-552.

9. Domling A (2008) Small molecular weight protein-protein interaction antagonists: an insurmountable challenge? *Current Opinion in Chemical Biology* 12: 281-291.

10. Pagliaro L, Felding J, Audouze K, Nielsen SJ, Terry RB, Krog-Jensen C, Butcher S (2004) Emerging classes of protein-protein interaction inhibitors and new tools for their development. *Current Opinion in Chemical Biology* 8: 442-449.

11. Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. *Science* 267: 383-386.

12. Atwell S, Ultsch M, De Vos AM, Wells JA (1997) Structural plasticity in a remodeled protein-protein interface. *Science* 278: 1125-1128.

13. Arkin MR, Randal M, DeLano WL, Hyde J, Luong TN, Oslob JD, Raphael DR, Taylor L, Wang J, McDowell RS, Wells JA, Braisted AC (2003) Binding of small molecules to an adaptive protein-protein interface. *Proc Natl Acad Sci USA* 100: 1603-1608.

14. Spring DR (2005) Chemical genetics to chemical genomics: small molecules offer big insights. *Chem Soc Rev* 34: 472-482.

15. Crews CM, Splittgerber U (1999) Chemical genetics: exploring and controlling cellular processes with chemical probes. *Trends Biochem Sci* 24: 317-320.

16. Fernandez PC, Frank SR, Wang L, Schroeder M, Liu S, Greene J, Cocito A, Amati B (2003) Genomic targets of the human c-Myc protein. *Genes Dev* 17: 1115-1129.

17. Levens DL (2003) Reconstructing MYC. *Genes Dev* 17: 1071-1077.

18. Baudino TA, Cleveland JL (2001) The Max network gone mad. *Mol Cell Biol* 21: 691-702.

19. Grandori C, Cowley SM, James LP, Eisenman RN (2000) The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu Rev Cell Dev Biol* 16: 653-699.

20. Henriksson M, Luscher B (1996) Proteins of the Myc network: essential regulators of cell growth and differentiation. *Adv Cancer Res* 68: 109-182.

21. Luscher B, Larsson LG (1999) The basic region/helix-loop-helix/leucine zipper domain of Myc proto-oncoproteins: function and regulation. *Oncogene* 18: 2955-2966.

22. Oster SK, Ho CS, Soucie EL, Penn LZ (2002) The myc oncogene: MarvelouslY Complex. *Adv Cancer Res* 84: 81-154.

23. Nesbit CE, Tersak JM, Prochownik EV (1999) MYC oncogenes and human neoplastic disease. *Oncogene* 18: 3004-3016.

24. Jenkins RB, Qian J, Lieber MM, Bostwick DG (1997) Detection of c-myc oncogene amplification and chromosomal anomalies in metastatic prostatic carcinoma by fluorescence in situ hybridization. *Cancer Res* 57: 524-531.

25. Prochownik EV (2004) c-Myc as a therapeutic target in cancer. *Expert Rev Anticancer Ther* 4: 289-302.

26. Yin X, Giap C, Lazo JS, Prochownik EV (2003) Low molecular weight inhibitors of Myc-Max interaction and function. *Oncogene* 22: 6151-6159.

27. Wang H, Hammoudeh DI, Follis AV, Reese BE, Lazo JS, Metallo SJ, Prochownik EV (2007) Improved low molecular weight Myc-Max inhibitors. *Mol Cancer Ther* 6: 2399-2408.

28. Hammoudeh DI, Follis AV, Prochownik EV, Metallo SJ (2009) Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J Am Chem Soc* 131: 7390-7401.

29. Mustata G, Follis AV, Hammoudeh DI, Metallo SJ, Wang H, Prochownik EV, Lazo JS, Bahar I (2009) Discovery of novel myc-max heterodimer disruptors with a three-dimensional pharmacophore model. *Journal of Medicinal Chemistry* 52: 1247-1250.

30. Jiang H, Bower KE, Beuscher AE, Zhou B, Bobkov AA, Olson AJ, Vogt PK (2009) Stabilizers of the Max homodimer identified in virtual ligand screening inhibit Myc function. *Molecular Pharmacology* 76: 491-502.

31. Meireles LM, Mustata G (2011) Discovery of modulators of protein-protein interactions: current approaches and limitations. *Curr Top Med Chem* 11: 248-257.

32. Hüser, J. (2006) High-throughput screening in drug discovery. Weinheim: Wiley-VCH.

33. Golebiowski A, Klopfenstein SR, Portlock DE (2003) Lead compounds discovered from libraries: part 2. *Current Opinion in Chemical Biology* 7: 308-325.

34. Macarron R (2006) Critical review of the role of HTS in drug discovery. *Drug Discovery Today* 11: 277-279.

35. Spencer RW (1998) High-throughput screening of historic collections: observations on file size, biological targets, and file diversity. *Biotechnology and Bioengineering* 61: 61-67.

36. Bajorath J (2002) Integration of virtual and high-throughput screening. *Nature Reviews Drug Discovery* 1: 882-894.

37. Haney SA, LaPan P, Pan J, Zhang J (2006) High-content screening moves to the front of the line. *Drug Discovery Today* 11: 889-894.

38. Ciruela F (2008) Fluorescence-based methods in the study of protein-protein interactions in living cells. *Current Opinion in Biotechnology* 19: 338-343.

39. Lundholt BK, Heydorn A, Bjorn SP, Praestegaard M (2006) A simple cell-based HTS assay system to screen for inhibitors of p53- Hdm2 protein-protein interactions. *Assay and Drug Development Technologies* 4: 679-688.

40. Paulmurugan R, Gambhir SS (2005) Novel fusion protein approach for efficient high-throughput screening of small molecule-mediating protein-protein interactions in cells and living animals. *Cancer Research* 65: 7413-7420.

41. Colas P (2008) High-throughput screening assays to discover small-molecule inhibitors of protein interactions. *Current Drug Discovery Technologies* 5: 190-199.

42. Fletcher S, Hamilton AD (2007) Protein-protein interaction inhibitors: small molecules from screening techniques. *Current Topics in Medicinal Chemistry* 7: 922-927.

43. Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, Kong N, Kammlott U, Lukacs C, Klein C, Fotouhi N, Liu EA (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303: 844-848.

44. Grasberger BL, Lu T, Schubert C, Parks DJ, Carver TE, Koblish HK, Cummings MD, LaFrance LV, Milkiewicz KL, Calvo RR, Maguire D, Lattanze J, Franks CF, Zhao S, Ramachandren K, Bylebyl GR, Zhang M, Manthey CL, Petrella EC, Pantoliano MW, Deckman IC, Spurlino JC, Maroney AC, Tomczuk BE, Molloy CJ, Bone RF (2005) Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. *Journal of Medicinal Chemistry* 48: 909-912.

45. Chene P (2003) Inhibiting the p53-MDM2 interaction: an important target for cancer therapy. *Nature Reviews Cancer* 3: 102-109.

46. Haupt Y, Maya R, Kazaz A, Oren M (1997) Mdm2 promotes the rapid degradation of p53. *Nature* 387: 296-299.

47. Coyne AG, Scott DE, Abell C (2010) Drugging challenging targets using fragment-based approaches. *Curr Opin Chem Biol* .

48. Murray CW, Rees DC (2009) The rise of fragment-based drug discovery. *Nature Chemistry* 1: 187-192.

49. de Kloe GE, Bailey D, Leurs R, de Esch IJ (2009) Transforming fragments into candidates: small becomes big in medicinal chemistry. *Drug Discovery Today* 14: 630-646.

50. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent developments in fragment-based drug discovery. *Journal of Medicinal Chemistry* 51: 3661-3680.

51. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nature Reviews Drug Discovery* 6: 211-219.

52. Erlanson DA (2006) Fragment-based lead discovery: a chemical update. *Current Opinion in Biotechnology* 17: 643-652.

53. Carr RA, Congreve M, Murray CW, Rees DC (2005) Fragment-based lead discovery: leads by design. *Drug Discovery Today* 10: 987-992.

54. Rees DC, Congreve M, Murray CW, Carr R (2004) Fragment-based lead discovery. *Nature Reviews Drug Discovery* 3: 660-672.

55. Erlanson DA, McDowell RS, O'Brien T (2004) Fragment-based drug discovery. *Journal of Medicinal Chemistry* 47: 3463-3482.

56. Schade M, Oschkinat H (2005) NMR fragment screening: tackling protein-protein interaction targets. *Current Opinion in Drug Discovery and Development* 8: 365-373.

57. Jhoti H, Cleasby A, Verdonk M, Williams G (2007) Fragment-based screening using X-ray crystallography and NMR spectroscopy. *Current Opinion in Chemical Biology* 11: 485-493.

58. Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H (2005) Fragment-based lead discovery using X-ray crystallography. *Journal of Medicinal Chemistry* 48: 403-413.

59. Hesterkamp T, Barker J, Davenport A, Whittaker M (2007) Fragment based drug discovery using fluorescence correlation: spectroscopy techniques: challenges and solutions. *Current Topics in Medicinal Chemistry* 7: 1582-1591.

60. Barker J, Courtney S, Hesterkamp T, Ullmann D, Whittaker M (2006) Fragment screening by biochemical assay. *Expert Opin Drug Discovery* 1: 225-236.

61. Erlanson DA, Wells JA, Braisted AC (2004) Tethering: fragment-based drug discovery. *Annu Rev Biophys Biomol Struct* 33: 199-223.

62. Hann MM, Leach AR, Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *Journal of Chemical Information and Computer Sciences* 41: 856-864.

63. McGovern SL, Caselli E, Grigorieff N, Shoichet BK (2002) A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *Journal of Medicinal Chemistry* 45: 1712-1722.

64. Bembenek SD, Tounge BA, Reynolds CH (2009) Ligand efficiency and fragment-based drug discovery. *Drug Discovery Today* 14: 278-283.

65. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* 9: 430-431.

66. Oltersdorf T, Elmore SW, Shoemaker AR, Armstrong RC, Augeri DJ, Belli BA, Bruncko M, Deckwerth TL, Dinges J, Hajduk PJ, Joseph MK, Kitada S, Korsmeyer SJ, Kunzer AR, Letai A, Li C, Mitten MJ, Nettesheim DG, Ng S, Nimmer PM, O'Connor JM, Oleksijew A, Petros AM, Reed JC, Shen W, Tahir SK, Thompson CB, Tomaselli KJ, Wang B, Wendt MD, Zhang H, Fesik SW, Rosenberg SH (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 435: 677-681.

67. Kirkin V, Joos S, Zornig M (2004) The role of Bcl-2 family members in tumorigenesis. *Biochimica et Biophysica Acta* 1644: 229-249.

68. Adams JM, Cory S (1998) The Bcl-2 protein family: arbiters of cell survival. *Science* 281: 1322-1326.

69. Petros AM, Dinges J, Augeri DJ, Baumeister SA, Betebenner DA, Bures MG, Elmore SW, Hajduk PJ, Joseph MK, Landis SK, Nettesheim DG, Rosenberg SH, Shen W, Thomas S, Wang X, Zanze I, Zhang H, Fesik SW (2006) Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis. *Journal of Medicinal Chemistry* 49: 656-663.

70. Wendt MD, Shen W, Kunzer A, McClellan WJ, Bruncko M, Oost TK, Ding H, Joseph MK, Zhang H, Nimmer PM, Ng SC, Shoemaker AR, Petros AM, Oleksijew A, Marsh K, Bauch J, Oltersdorf T, Belli BA, Martineau D, Fesik SW, Rosenberg SH, Elmore SW (2006) Discovery and structure-activity relationship of antagonists of B-cell lymphoma 2 family proteins with chemopotentiation activity in vitro and in vivo. *Journal of Medicinal Chemistry* 49: 1165-1181.

71. Bruncko M, Oost TK, Belli BA, Ding H, Joseph MK, Kunzer A, Martineau D, McClellan WJ, Mitten M, Ng SC, Nimmer PM, Oltersdorf T, Park CM, Petros AM, Shoemaker AR, Song X, Wang X, Wendt MD, Zhang H, Fesik SW, Rosenberg SH, Elmore SW (2007) Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL. *Journal of Medicinal Chemistry* 50: 641-662.

72. Park CM, Bruncko M, Adickes J, Bauch J, Ding H, Kunzer A, Marsh KC, Nimmer P, Shoemaker AR, Song X, Tahir SK, Tse C, Wang X, Wendt MD, Yang X, Zhang H, Fesik SW, Rosenberg SH, Elmore SW (2008) Discovery of an orally bioavailable small molecule inhibitor of prosurvival B-cell lymphoma 2 proteins. *Journal of Medicinal Chemistry* 51: 6902-6915.

73. ClinicalTrials.gov A Phase 1/2a Study of ABT-263 in Subjects With Small Cell Lung Cancer (SCLC) or Other Non-Hematological Malignancies. Available: http://clinicaltrials.gov/show/NCT00445198.

74. Gron H, Hyde-DeRuyscher R (2000) Peptides as tools in drug discovery. *Current Opinion in Drug Discovery and Development* 3: 636-645.

75. Ladner RC, Sato AK, Gorzelany J, de Souza M (2004) Phage display-derived peptides as therapeutic alternatives to antibodies. *Drug Discovery Today* 9: 525-529.

76. Sato AK, Viswanathan M, Kent RB, Wood CR (2006) Therapeutic peptides: technological advances driving peptides into development. *Current Opinion in Biotechnology* 17: 638-642.

77. FitzGerald K (2000) In vitro display technologies - new tools for drug discovery. *Drug Discovery Today* 5: 253-258.

78. Amstutz P, Forrer P, Zahnd C, Pluckthun A (2001) In vitro display technologies: novel developments and applications. *Current Opinion in Biotechnology* 12: 400-405.

79. Rothe A, Hosse RJ, Power BE (2006) In vitro display technologies reveal novel biopharmaceutics. *FASEB Journal* 20: 1599-1610.

80. Kay BK, Kurakin AV, Hyde-DeRuyscher R (1998) From peptides to drugs via phage display. *Drug Discovery Today* 3: 370-378.

81. Cunningham BC, Wells JA (1997) Minimized proteins. *Curr Opin Struct Biol* 7: 457-462.

82. Sidhu SS, Fairbrother WJ, Deshayes K (2003) Exploring protein-protein interactions with phage display. *ChemBioChem* 4: 14-25.

83. Ahn JM, Boyle NA, MacDonald MT, Janda KD (2002) Peptidomimetics and peptide backbone modifications. *Mini Reviews in Medicinal Chemistry* 2: 463-473.

84. Patch JA, Barron AE (2002) Mimicry of bioactive peptides via non-natural, sequence-specific peptidomimetic oligomers. *Current Opinion in Chemical Biology* 6: 872-877.

85. Kieber-Emmons T, Murali R, Greene MI (1997) Therapeutic peptides and peptidomimetics. *Current Opinion in Biotechnology* 8: 435-441.

86. Adessi C, Soto C (2002) Converting a peptide into a drug: strategies to improve stability and bioavailability. *Current Medicinal Chemistry* 9: 963-978.

87. Hamman JH, Enslin GM, Kotze AF (2005) Oral delivery of peptide drugs: barriers and developments. *BioDrugs* 19: 165-177.

88. Morishita M, Peppas NA (2006) Is the oral route possible for peptide and protein drug delivery? *Drug Discovery Today* 11: 905-910.

89. Salvesen GS, Duckett CS (2002) IAP proteins: blocking the road to death's door. *Nature Reviews Molecular Cell Biology* 3: 401-410.

90. Liu Z, Sun C, Olejniczak ET, Meadows RP, Betz SF, Oost T, Herrmann J, Wu JC, Fesik SW (2000) Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. *Nature* 408: 1004-1008.

91. Wu G, Chai J, Suber TL, Wu JW, Du C, Wang X, Shi Y (2000) Structural basis of IAP recognition by Smac/DIABLO. *Nature* 408: 1008-1012.

92. Du C, Fang M, Li Y, Li L, Wang X (2000) Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition. *Cell* 102: 33-42.

93. Sun H, Nikolovska-Coleska Z, Yang CY, Qian D, Lu J, Qiu S, Bai L, Peng Y, Cai Q, Wang S (2008) Design of small-molecule peptidic and nonpeptidic Smac mimetics. *Accounts of Chemical Research* 41: 1264-1277.

94. Sun W, Nikolovska-Coleska Z, Qin D, Sun H, Yang CY, Bai L, Qiu S, Wang Y, Ma D, Wang S (2009) Design, synthesis, and evaluation of potent, nonpeptidic mimetics of second mitochondria-derived activator of caspases. *Journal of Medicinal Chemistry* 52: 593-596.

95. Ascenta Therapeutics. Available: http://www.ascenta.com.

96. Souroujon MC, Mochly-Rosen D (1998) Peptide modulators of protein-protein interactions in intracellular signaling. *Nature Biotechnology* 16: 919-924.

97. Ducki S, Bennett E (2009) Protein-protein interactions: recent progress in the development of selective PDZ inhibitors. *Current Chemical Biology* 3: 146-158.

98. Sawyer TK, Bohacek RS, Dalgarno DC, Eyermann CJ, Kawahata N, Metcalf CA, III, Shakespeare WC, Sundaramoorthi R, Wang Y, Yang MG (2002) SRC homology-2 inhibitors: peptidomimetic and nonpeptide. *Mini Reviews in Medicinal Chemistry* 2: 475-488.

99. Witthuhn BA, Quelle FW, Silvennoinen O, Yi T, Tang B, Miura O, Ihle JN (1993) JAK2 associates with the erythropoietin receptor and is tyrosine phosphorylated and activated following stimulation with erythropoietin. *Cell* 74: 227-236.

100. Wrighton NC, Farrell FX, Chang R, Kashyap AK, Barbone FP, Mulcahy LS, Johnson DL, Barrett RW, Jolliffe LK, Dower WJ (1996) Small peptides as potent mimetics of the protein hormone erythropoietin. *Science* 273: 458-464.

101. Livnah O, Stura EA, Johnson DL, Middleton SA, Mulcahy LS, Wrighton NC, Dower WJ, Jolliffe LK, Wilson IA (1996) Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 A. *Science* 273: 464-471.

102. Johnson DL, Farrell FX, Barbone FP, McMahon FJ, Tullai J, Hoey K, Livnah O, Wrighton NC, Middleton SA, Loughney DA, Stura EA, Dower WJ, Mulcahy LS, Wilson IA, Jolliffe LK (1998) Identification of a 13 amino acid peptide mimetic of erythropoietin and description of amino acids critical for the mimetic activity of EMP1. *Biochemistry* 37: 3699-3710.

103. Livnah O, Johnson DL, Stura EA, Farrell FX, Barbone FP, You Y, Liu KD, Goldsmith MA, He W, Krause CD, Pestka S, Jolliffe LK, Wilson IA (1998) An antagonist peptide-EPO receptor complex suggests that receptor dimerization is not sufficient for activation. *Nature Structural Biology* 5: 993-1004.

104. Beeley NR (2000) Can peptides be mimicked? *Drug Discovery Today* 5: 354-363.

105. Jiang G, Hunter T (1999) Receptor signaling: when dimerization is not enough. *Current Biology* 9: R568-R571.

106. Macdougall IC (2008) Hematide, a novel peptide-based erythropoiesis-stimulating agent for the treatment of anemia. *Current Opinion in Investigational Drugs* 9: 1034-1047.

107. Marshall GR, Kuster DJ, Che Y (2009) Chemogenomics with protein secondary-structure mimetics. *Methods Mol Biol* 575: 123-158.

108. Hershberger SJ, Lee SG, Chmielewski J (2007) Scaffolds for blocking protein-protein interactions. *Current Topics in Medicinal Chemistry* 7: 928-942.

109. Orner BP, Ernst JT, Hamilton AD (2001) Toward proteomimetics: terphenyl derivatives as structural and functional mimics of extended regions of an alpha-helix. *J Am Chem Soc* 123: 5382-5383.

110. Ernst JT, Becerril J, Park HS, Yin H, Hamilton AD (2003) Design and application of an alpha-helix-mimetic scaffold based on an oligoamide-foldamer strategy: antagonism of the Bak BH3/Bcl-xL complex. *Angewandte Chemie International Edition in English* 42: 535-539.

111. Yin H, Hamilton AD (2004) Terephthalamide derivatives as mimetics of the helical region of Bak peptide target Bcl-xL protein. *Bioorganic & Medicinal Chemistry Letters* 14: 1375-1379.

112. Antuch W, Menon S, Chen QZ, Lu Y, Sakamuri S, Beck B, Schauer-Vukasinovic V, Agarwal S, Hess S, Domling A (2006) Design and modular parallel synthesis of a MCR derived alpha-helix mimetic protein-protein interaction inhibitor scaffold. *Bioorganic & Medicinal Chemistry Letters* 16: 1740-1743.

113. Shaginian A, Whitby LR, Hong S, Hwang I, Farooqi B, Searcey M, Chen J, Vogt PK, Boger DL (2009) Design, Synthesis, and Evaluation of an alpha-Helix Mimetic Library Targeting Protein-Protein Interactions. *J Am Chem Soc* 131: 5564-5572.

114. Kutzki O, Park HS, Ernst JT, Orner BP, Yin H, Hamilton AD (2002) Development of a potent Bcl-x(L) antagonist based on alpha-helix mimicry. *J Am Chem Soc* 124: 11838-11839.

115. Klebe G (2000) Recent developments in structure-based drug design. *Journal of Molecular Medicine* 78: 269-281.

116. Anderson AC (2003) The process of structure-based drug design. *Chemistry & Biology* 10: 787-797.

117. Gane PJ, Dean PM (2000) Recent advances in structure-based rational drug design. *Curr Opin Struct Biol* 10: 401-404.

118. Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M (2004) Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J Med Chem* 47: 45-55.

119. Stahura FL, Bajorath M (2005) New methodologies for ligand-based virtual screening. *Current Pharmaceutical Design* 11: 1189-1202.

120. Johnson, M. A. and Maggiora, G. M. (1990) Concepts and applications of molecular similarity. New York: John Wiley & Sons, Inc.

121. Rognan D (2007) Chemogenomic approaches to rational drug design. *British Journal of Pharmacology* 152: 38-52.

122. Sheridan RP, Kearsley SK (2002) Why do we need so many chemical similarity search methods? *Drug Discovery Today* 7: 903-911.

123. Hert J, Willett P, Wilton DJ, Acklin P, Azzaoui K, Jacoby E, Schuffenhauer A (2004) Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures. *Journal of Chemical Information and Computer Sciences* 44: 1177-1185.

124. Meireles L, Gur M, Bakan A, Bahar I (in press) Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins. *Protein Science*.

125. Bakan A, Bahar I (2009) The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc Natl Acad Sci USA* 106: 14349-14354.

126. Bahar I, Lezon TR, Yang LW, Eyal E (2010) Global dynamics of proteins: bridging between structure and function. *Annu Rev Biophys* 39: 23-42.

127. Fischer E (1894) Einfluss der configuration auf die wirkung der enzyme. *Ber Dtsch Chem Ges* 27: 2985-2993.

128. Koshland DE (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA* 44: 98-104.

129. Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: a plausible model. *J Mol Biol* 12: 88-118.

130. Boehr DD, Nussinov R, Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* 5: 789-796.

131. Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett* 77: 1905-1908.

132. Ma JP (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* 13: 373-380.

133. Bahar I, Chennubhotla C, Tobi D (2007) Intrinsic dynamics of enzymes in the unbound state and, relation to allosteric regulation. *Curr Opin Struct Biol* 17: 633-640.

134. Tama F, Brooks CL (2006) Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annu Rev Biophys Biomol Struct* 35: 115-133.

135. Yang LW, Chng CP (2008) Coarse-grained models reveal functional dynamics--I. Elastic network models--theories, comparisons and perspectives. *Bioinform Biol Insights* 2: 25-45.

136. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 2: 173-181.

137. Haliloglu T, Bahar I, Erman B (1997) Gaussian dynamics of folded proteins. *Phys Rev Lett* 79: 3090-3093.

138. Doruker P, Atilgan AR, Bahar I (2000) Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. *Proteins* 40: 512-524.

139. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 80: 505-515.

140. Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. *Protein Eng* 14: 1-6.

141. Flory PJ (1976) Statistical thermodynamics of random networks. *Proc R Soc London,A* 351: 351-380.

142. Hinsen K (1998) Analysis of domain motions by approximate normal mode calculations. *Proteins:Struct, Funct, Genet* 33: 417-429.

143. Hinsen K (2006) Normal mode theory and harmonic potential approximations. In: Cui Q, Bahar I, editors. Normal mode analysis: theory and applications to biological and chemical systems. Boca Raton, FL: Chapman & Hall/CRC. pp. 1-16.

144. Yang L, Song G, Jernigan RL (2009) Protein elastic network models and the ranges of cooperativity. *Proc Natl Acad Sci USA* 106: 12347-12352.

145. Riccardi D, Cui Q, Phillips GN (2009) Application of elastic network models to proteins in the crystalline state. *Biophys J* 96: 464-475.

146. Hinsen K, Petrescu AJ, Dellerue S, Bellissent-Funel MC, Kneller GR (2000) Harmonicity in slow protein dynamics. *Chem Phys* 261: 25-37.

147. Kovacs JA, Chacon P, Abagyan R (2004) Predictions of protein flexibility: first-order measures. *Proteins:Struct, Funct, Bioinf* 56: 661-668.

148. Lezon TR, Bahar I (2010) Using entropy maximization to understand the determinants of structural dynamics beyond native contact topology. *PLoS Comput Biol* 6: e1000816.

149. Xu CY, Tobi D, Bahar I (2003) Allosteric changes in protein structure computed by a simple mechanical model: Hemoglobin T <-> R2 transition. *J Mol Biol* 333: 153-168.

150. Yang L, Song G, Carriquiry A, Jernigan RL (2008) Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure* 16: 321-330.

151. Tobi D, Bahar I (2005) Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci USA* 102: 18908-18913.

152. Keskin O (2007) Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies. *BMC Struct Biol* 7.

153. Lange OF, Lakomek NA, Fares C, Schroder GF, Walter KFA, Becker S, Meiler J, Grubmuller H, Griesinger C, de Groot BL (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320: 1471-1475.

154. Gsponer J, Christodoulou J, Cavalli A, Bui JM, Richter B, Dobson CM, Vendruscolo M (2008) A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction. *Structure* 16: 736-746.

155. Bakan A, Meireles LM, Bahar I (2011) ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics* 27: 1575-1577.

156. Totrov M, Abagyan R (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol* 18: 178-184.

157. Eyal E, Yang LW, Bahar I (2006) Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* 22: 2619-2627.

158. Suhre K, Sanejouand YH (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res* 32: W610-W614.

159. Camps J, Carrillo O, Emperador A, Orellana L, Hospital A, Rueda M, Cicin-Sain D, D'Abramo M, Gelpi JL, Orozco M (2009) FlexServ: an integrated tool for the analysis of protein flexibility. *Bioinformatics* 25: 1709-1710.

160. Mashiach E, Nussinov R, Wolfson HJ (2010) FiberDock: a web server for flexible induced-fit backbone refinement in molecular docking. *Nucleic Acids Res* 38: W457-W461.

161. Cavasotto CN, Kovacs JA, Abagyan RA (2005) Representing receptor flexibility in ligand docking through relevant normal modes. *J Am Chem Soc* 127: 9632-9640.

162. Mashiach E, Nussinov R, Wolfson HJ (2010) FiberDock: Flexible induced-fit backbone refinement in molecular docking. *Proteins* 78: 1503-1519.

163. Rueda M, Bottegoni G, Abagyan R (2009) Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes. *J Chem Inf Model* 49: 716-725.

164. Floquet N, Marechal JD, Badet-Denisot MA, Robert CH, Dauchez M, Perahia D (2006) Normal mode analysis as a prerequisite for drug design: application to matrix metalloproteinases inhibitors. *FEBS Lett* 580: 5130-5136.

165. Sperandio O, Mouawad L, Pinto E, Villoutreix BO, Perahia D, Miteva MA (2010) How to choose relevant multiple receptor conformations for virtual screening: a test case of Cdk2 and normal mode analysis. *EUR BIOPHYS J BIOPHY* 39: 1365-1372.

166. Zacharias M, Sklenar H (1999) Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: Application to DNA minor groove ligand complex. *J Comput Chem* 20: 287-300.

167. Zacharias M (2004) Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: binding of FK506 to FKBP. *Proteins* 54: 759-767.

168. May A, Zacharias M (2007) Protein-protein docking in CAPRI using ATTRACT to account for global and local flexibility. *Proteins:Struct, Funct, Bioinf* 69: 774-780.

169. May A, Zacharias M (2008) Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking. *Proteins:Struct, Funct, Bioinf* 70: 794-809.

170. May A, Zacharias M (2008) Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: evaluation on kinase inhibitor cross docking. *J Med Chem* 51: 3499-3506.

171. Lindahl E, Delarue M (2005) Refinement of docked protein-ligand and protein-DNA structures using low frequency normal mode amplitude optimization. *Nucleic Acids Res* 33: 4496-4506.

172. Kingsford CL, Chazelle B, Singh M (2005) Solving and analyzing side-chain positioning problems using linear and integer programming. *Bioinformatics* 21: 1028-1036.

173. Desmet J, De MM, Hazes B, Lasters I (1992) The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* 356: 539-542.

174. Gurobi Optimization. Available: http://www.gurobi.com.

175. IBM ILOG CPLEX Optimizer. Available: http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/.

176. GLPK: GNU Linear Programming Kit. Available: http://www.gnu.org/software/glpk/.

177. Grant JA, Pickup BT (1995) A Gaussian description of molecular shape. *Journal of Physical Chemistry* 99: 3503-3510.

178. Grant JA, Gallardo MA, Pickup BT (1996) A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J Comput Chem* 17: 1653-1666.

179. Grant JA, Pickup BT (1997) Gaussian shape methods. In: Computer Simulation of Biomolecular Systems.

180. Bostrom J, Greenwood JR, Gottfries J (2003) Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J Mol Graph Model* 21: 449-462.

181. University of Pittsburgh Center for Chemical Methodologies and Library Development. Available: http://ccc.chem.pitt.edu/UPCMLD/.

182. Maass P, Schulz-Gasch T, Stahl M, Rarey M (2007) Recore: a fast and versatile method for scaffold hopping based on small molecule crystal structure conformations. *J Chem Inf Model* 47: 390-399.

183. Feig M, Rotkiewicz P, Kolinski A, Skolnick J, Brooks CL, III (2000) Accurate reconstruction of all-atom protein representations from side-chain-based low-resolution models. *Proteins* 41: 86-97.

184. Heath AP, Kavraki LE, Clementi C (2007) From coarse-grain to all-atom: toward multiscale analysis of protein landscapes. *Proteins* 68: 646-661.

185. Rotkiewicz P, Skolnick J (2008) Fast procedure for reconstruction of full-atom protein models from reduced representations. *J Comput Chem* 29: 1460-1465.