Computational, Synthetic, Biochemical and Biological Studies and Characterization on
STAT3 Inhibitors for Potential Anticancer Therapy

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University

By
Wenying Yu, M.S.
Graduate Program in Pharmacy

The Ohio State University
2013

Dissertation Committee:
Chenglong Li, Advisor
Karl Werbovetz,
Werner Tjarks
Copyright by

Wenying Yu

2013
Abstract

The objective of this dissertation is to design and discover promising homo-dimerization inhibitors on Signal Transducer and Activator of Transcription 3 (STAT3) protein for anticancer therapy. Constitutive activation of STAT3 has been found in nearly every major cancer, including solid tumors, blood tumors and lymphoma, specific cancer types including breast cancers, pancreatic cancers, sarcoma cancers, etc. It has been validated as an attractive therapeutic target for cancer therapy. To block both STAT3 activation and dimerization, a viable strategy is to design inhibitors competing with the native phosphotyrosine peptide that binds to the STAT3 SH2 domain.

The work can be divided into four stages including computational, synthetic, biochemical and biological studies. For the computational study, an improved fragment-based drug design (FBDD) strategy, in silico site-directed FBDD, was applied in this study to discover novel series of STAT3 inhibitors. Multiple Ligand Simultaneous Docking (MLSD) as an alternative method was also used to search for inhibitors which can tightly bind to STAT3 SH2 domain. For the synthetic study, we found the first copper-catalyzed one-pot highly regioselective C-N coupling of substituted naphthoquinones and chemoselective intramolecular ring fusion of sulfonamide synthetic approaches. Facile chemoselective synthetic routes were discovered to obtain diverse ring-opening and
reversible ring-fused compounds with great functional group tolerance. Stereochemistry was confirmed by NMRs. For the biochemical study, the binding between 5,8-dioxo-6-(pyridin-3-ylamino)-5,8-dihydronaphthalene-1-sulfonamide (LY5) and STAT3 SH2 domain was confirmed by fluorescence polarization assay. The binding between 5,8-dioxo-6-(3-(piperazin-1-yl)phenylamino)-5,8-dihydronaphthalene-1-sulfonamide (LY17) and STAT3 SH2 domain was confirmed by microscale thermophoresis (MST) tests. For the biological assays, four out of the five selected and synthesized compounds from in silico site-directed FBDD method have IC\textsubscript{50} values lower than 5 µM for the U2OS cancer cells. LY5 has an IC\textsubscript{50} range in 0.5-1.4 µM in various cancer cell lines. Using the MLSD method, LY17 has an IC\textsubscript{50} about 1.15 µM and LY13 has an IC\textsubscript{50} about 1.49 µM in UW426 cancer cell line. For the in vivo study, LY5 was also found significantly suppressed tumor growth in mice. The results have demonstrated the feasibility of both in silico site-directed FBDD and MLSD approaches for STAT3 drug discovery, which can be applied in other drug targets in general too. LY5 and LY17 were found to be very promising drug candidates and their druggability will be confirmed by further studies.
Dedication

This document is dedicated to my husband Yang Li and my parents.
Acknowledgments

This five-year PhD training not only increases my age but also lets me meet so many wonderful and helpful people that I would like to express my deeply felt gratitude to.

First of all, I would like to appreciate my mentor, Dr. Chenglong Li, for his all-aspect unlimited support on both my research and my life in OSU. There is a saying in China, “A teacher for a day is a father for a lifetime.” I greatly appreciate his fatherly guidance to lead me into the area of computational drug design and high expectations to encourage me to continue learning. I would like to express my gratitude to Dr. Li, for his great inspiration and warm encouragement whenever I needed. His insights and vast wisdom in many interdisciplines of computational drug design, organic synthesis, biochemical and biological assays also greatly broadened my research vision.

I am also grateful to Dr. Tjarks and Dr. Werbovetz for their great help in my course study, research, and my TA work. I also deeply appreciate their careful proofreading on my candidacy proposal. I would like to thank Dr. Jiayuh Lin for his guidance and support during my research at the nationwide children’s hospital.
I should express my gratitude to Dr. Fuchs and Dr. Kinghorn for their kindly support in my research work. I should especially appreciate Ryan E. Pavlovicz, Dr. Xiaohua Zhu and Dr. Sihui Long for all their great help in my research and in my paper editing. I should thank our previous lab members, Dr. In-Hee Park and Dr. Kiran Mahasenan, and Dr. Huameng Li for their generous help in my computational studies. The discussion with Dr. Sihui Long and Dr. Ribai Yan about my synthesis study always assisted me to solve all kinds of problems. I should appreciate Dr. Xiao Hui, Veronica Olson, Dr. Aiguo Liu, Dr. Li Lin and all the other friends at the nationwide children’s hospital for their help in my biological studies. I should thank Dr. Deng Ye and Li Pan, the Couple, for all their help in both my daily life and my research study. I am also very grateful to Jonathan P. Etter for his assistance in FP assays. I’d like to express my gratitude to all my previous and current lab-mates, for their generous help, interesting scientific discussions and fun time that we experienced together. I will always miss being around with so many wonderful people.

Last but not the least, I would like to thank my parents, without their selfless support, I would have never made it this far. I’d like to thank my dear sister for taking care of our parents and letting me focus on my research. And finally, I am so grateful to my husband Yang for always considering my needs first and for always standing by my side. Thank you, my love.
Vita

2002 .........................................................B.S. National Life science and technology
   Base for Talented Persons, China
   Pharmaceutical University
2006 ..........................................................M.S. Medicinal Chemistry, China
   Pharmaceutical University
2008 to present .........................................Graduate Teaching Associate, Department
   of Medicinal Chemistry, School of
   Pharmacy, The Ohio State University

Publications

1. Wenying Yu, Hui Xiao, Jiayuh Lin, Chenglong Li (2013). Discovery of novel STAT3
   small molecule inhibitors via in silico site-directed fragment-based drug design. J. Med.
   Chem. 56, 4402-4412.

2. Wenying Yu, Chenglong Li (2013) 4-nitrophenyl 2-sulfamoylphenylcarbamate. Acta

3. Wenying Yu, Chenglong Li. Regioselective one-pot C-N coupling of substituted
   naphthoquinones: selective intramolecular ring fusion of sulfonamide. Tetrahedron
   (Under review)


**Fields of Study**

Major Field: Pharmacy (Pharmaceutical Sciences)
# Table of Contents

Abstract ................................................................................................................................. ii

Dedication ............................................................................................................................... iv

Acknowledgments .................................................................................................................. v

Vita ....................................................................................................................................... vii

List of Tables ........................................................................................................................ xiv

List of Figures ......................................................................................................................... xv

List of Schemes ....................................................................................................................... xix

Chapter 1: Computer aided drug design ............................................................................. 1

1.1 Introduction ..................................................................................................................... 1

1.2 Molecular docking and high throughput virtual screening ........................................... 2

1.3 Fragment-based drug design (FBDD) and in silico FBDD ........................................... 5

1.4 CombiGlide and Multiple Ligand Simultaneous Docking(MLSD) ....................... 12

1.5 Molecular Dynamics Simulation (MD) ........................................................................ 14

1.6 Induced-fit docking (IFD) ............................................................................................ 15

1.7 Free energy calculation ................................................................................................. 16

Chapter 2: Biochemical assays ........................................................................................... 18

2.1 Introduction .................................................................................................................... 18
5.3 Results .......................................................................................................................... 56

5.3.1 SAR study ................................................................................................................. 56

5.3.2 Biological and Biochemical Assays ........................................................................... 56

5.4 Discussion ..................................................................................................................... 57

5.5 Conclusion ................................................................................................................... 59

5.6 Experimental section ................................................................................................. 59

5.6.1 Chemistry .................................................................................................................. 59

5.6.2 Biological Assays. .................................................................................................... 59

Chapter 6: Computational design of STAT3 anticancer inhibitors using multiple ligand simultaneous docking method ......................................................................................... 61

6.1 Abstract ....................................................................................................................... 61

6.2 Introduction .................................................................................................................. 61

6.3 Results .......................................................................................................................... 62

6.3.1 Multifragment Simultaneous Docking Screening ....................................................... 64

6.3.2 Fragment Linkage ..................................................................................................... 64

6.3.3 Hit Selection .............................................................................................................. 65

6.3.4 Biological and Biochemical Assays .......................................................................... 67

6.4 Discussion ..................................................................................................................... 71

6.5 Conclusion ................................................................................................................... 73
6.6 Experimental section

6.6.1 MLSD .......................................................... 73

6.6.2 Biological Assays and biochemical assays........................................ 74

Chapter 7: Regioselective one-pot C-N coupling of substituted naphthoquinones:
selective intramolecular ring fusion of sulfonamide.............................. 77

7.1 Abstract ............................................................................. 77

7.2 Introduction ........................................................................ 77

7.3 Conclusions ....................................................................... 87

7.4 General Information ............................................................ 88

7.5 Materials .......................................................................... 88

7.6 The stereochemistry of the regioisomers ....................................... 91

7.7 Control experiments on the reaction mechanism ................................ 94

7.8 General procedure of C-N coupling reaction of ring-fused 6-amino/aniline-5H-
Naphth[1,8-cd] isothiazol-5-one,1,1-dioxides ........................................ 94

7.9 Characterization data of the ring-fused compounds........................ 95

7.10 General procedure of C-N coupling reaction of ring-opening 6-amino/aniline-
naphthalene-5,8-dione-1-sulfonamides .............................................. 99

7.11 Characterization data of the ring-opening compounds (3a-3g)........ 99

7.12 The hydrolysis reaction from the ring-fused compound to the ring-opening
compound.............................................................................. 103
Chapter 8: Dynamic induced-fit docking provides insight into the STAT3 inhibition. . 105

8.1 Abstract ............................................................................................................................................. 105

8.2 Introduction ...................................................................................................................................... 105

8.3 Motivation of this research ............................................................................................................. 106

8.4 Materials and Methods .................................................................................................................. 107

8.5 Conformational sampling of apo-STAT3 and holo-STAT3-ligand complex ...... 108

8.6 Extract representative structures of each MD run ........................................................................... 109

8.7 Ensemble Dockings .......................................................................................................................... 110

8.8 Structural analysis of the induced-fit structures of STAT3......................................................... 110

8.9 Results .............................................................................................................................................. 111

8.10 Discussion ....................................................................................................................................... 120

8.11 Conclusion ..................................................................................................................................... 120

Chapter 9: Summary and Conclusions ................................................................................................. 122

References............................................................................................................................................... 123

Appendix A. Copies $^1$H NMR, $^{13}$C NMR. .......................................................................................... 134

Appendix B. The docked fragment combinations by MLSD. ............................................................... 148

Appendix C. The reposition of selected compounds. ........................................................................... 151

Appendix D Computational drug design of natural products as STAT3 inhibitors ...... 153
List of Tables

Table 1.1 Comparisons of HTS and FBDD ................................................................. 7
Table 1.2 Potent inhibitors derived from FBDD \(^{12}\) ..................................................... 8
Table 1.3 The comparison of the screening methods .................................................... 10
Table 4.1 The comparison of conventional FBDD and in silico site-directed FBDD ...... 30
Table 4.2 Chemical structures, docking scores, and IC\(_{50}\) of the designed STAT3 inhibitors ................................................................................................................................. 37
Table 4.3 IC\(_{50}\) values for human sarcoma cells (RD2, RH30 and U2OS) inhibition by STAT3 inhibitors, 13 (LLL12), 8, and 11 .............................................................................................................. 38
Table 5.1 IC\(_{50}\) of the two series of STAT3 Inhibitors ...................................................... 57
Table 6.1 Predicted compounds and their selectivity to STAT3 and STAT1. ............... 66
Table 6.2 IC\(_{50}\) of designed STAT3 Inhibitors .................................................................. 68
Table 7.1 Optimization of reaction conditions ................................................................. 80
Table 8.1 The comparison of docking scores and IC\(_{50}\) values of the known STAT3 inhibitors and decoy compound .................................................................................................................... 112
Table 8.2 The Gscore difference between active compounds and LY10. ...................... 116
## List of Figures

Figure 1.1 The “lock-and-key” diagram represents a typical docking experiment. .......... 2

Figure 1.2 General procedures of FBDD and vHTS\(^9\) ........................................... 6

Figure 1.3 General strategies for evolving the fragments to larger hits\(^9\). ................. 11

Figure 3.1 The activation of STAT3 signaling pathway by intrinsic and extrinsic pathways\(^6\). .................................................................................................................. 22

Figure 3.2A. The Tyr705 residues of STAT3 monomers are phosphorylated then the homo-dimer binds to DNA, and finally cause cancer. B. The chemical structure of XZH-5. C. The binding mode of XZH-5 predicted by Autodock4. ........................................... 23

Figure 4.1 Docking model of STAT3 inhibitor 8 binding to the STAT3 SH2 domain (PDB: 1BG1), generated by AutoDock4 and viewed by Maestro. .............................. 26

Figure 4.2 Site-directed FBDD strategy ........................................................................... 28

Figure 4.3 Docking modes of the selected known STAT3 inhibitors with the STAT3 SH2 domain (PDB: 1BG1), generated by AutoDock4. Surface representation was created with Maestro. ................................................................. 32

Figure 4.4 Chemical structures of the selected STAT3 dimerization inhibitors.............. 33

Figure 4.5 Categorized fragment sub-libraries, site pTyr705 (1-7) and side pocket (a-g). (The fragment phosphotyrosine group was not included in the fragment libraries due to its peptidomimetic property.) .................................................................................. 34

Figure 4.6 The evolution of linker design............................................................................. 35
Figure 4.7 The docking modes of 13 (LLL12) and 8 (LY5). Both linkers, the hydroxyl group in 13 and secondary amine group in 8, form hydrogen bonds with the Ser636.

Figure 4.8 11 and 8 inhibit constitutive STAT3 phosphorylation.......................... 39

Figure 4.9 Compound 8 inhibits constitutive STAT3 phosphorylation.................. 40

Figure 4.10 11 and 8 inhibit IL-6 induced STAT3 phosphorylation in MCF-7 breast cancer cells........................................................................................................ 41

Figure 4.11 8 inhibits IL-6 induced STAT3 phosphorylation in Daoy cancer cells..... 42

Figure 4.12 IFN-γ induced STAT1 phosphorylation in Daoy cancer cells. ............. 43

Figure 4.13 Compound 8 suppresses tumor growth of MDA-MB-231 breast cancer-initiating cells in mouse tumor model in vivo................................................. 44

Figure 4.14 Compound 8 binds to STAT3 SH2 protein............................................ 45

Figure 5.1 The chemical structures of the cyclized and ring-opening STAT3 inhibitors. 55

Figure 6.1 Chemical structures of fragment 1 and LLL12. ...................................... 63

Figure 6.2 The docked conformations of fragments and linked compound LY17........ 65

Figure 6.3 LLL12, LY5 and predicted structures by MLSD. ..................................... 67

Figure 6.4 Human Medulloblastoma cell line UW426 was treated with LY17 (1-2.5 μM) and LY17-R2 (1-2.5 μM) for 12 h, then whole-cell extracts were prepared and phospho-STAT3 was detected by Western blot assay, revealing a decrease in STAT3 phosphorylation........................................................................................................ 69

Figure 6.5 LY17 and LY17-R2 inhibit IL-6 induced STAT3 phosphorylation in Daoy cancer cells. The Daoy cells were serum-starved overnight, then left untreated or treated
with LY17 or LY17-R2 (1-2.5 µM) for 2h, followed by stimulation by IL-6 (50 ng/mL). The cells were harvested at 30 minutes and analyzed by Western blot assays.

Figure 6.6 Microscale thermophoresis (MST) tests on LY17 as a modified STAT3 inhibitor.

Figure 7.1 Structures of naphthalene-1,4-dione derivatives.

Figure 7.2 Scope of ring-fused 6-amino/aniline-5H-naphth[1,8-cd]isothiazol-5-one,1,1-dioxides $^{ab}$.

Figure 7.3 Scope of ring-opening 6-amino/aniline-naphthalene-5,8-dione-1-sulfonamides $^{ac}$.

Figure 7.4 $^{13}$CNMR, DEPT-90 and HMBC of compound 3b.

Figure 7.5 $^1$H$^1$HCOSY of compound 3b.

Figure 7.6 HSQC of compound 3b.

Figure 8.1 Chemical structures of the test set compounds.

Figure 8.2 The RMSDs of the MD simulations of apo-STAT3 and holo-STAT3.

Figure 8.3 The $\Delta G_{\text{ligand-LY10}}$ surfaces of all the snapshots.

Figure 8.4 The comparison of residue-wise RMSDs of the selected induced-fit structures and the crystal structure (PDB:1BG1).

Figure 8.5 Two most fluctuant regions of protein, residues 622-627 and residues 661-665, are in two neighbor loops.

Figure 8.6 A parallel comparison of the crystal structure and the five selected induced-fit structures.

Figure B.1 Fragment 1 and polar-aromatic-amine fragments.
Figure B.2 Fragment 1 and polar-aromatic-amine fragments with hydroxyl substitution. ........................................................................................................................................................................... 149

Figure B.3 Fragment 1 and polar-aromatic-sulfonamide/polar-aromatic-cyclic amine fragments ........................................................................................................................................................................................................................................... 149

Figure B.4 Fragment 1 and polar-aromatic acid fragments ........................................................................................................................................................................................................................................... 150

Figure C.1 The reposition of LY9 to the docked fragments ........................................................................................................................................................................................................................................... 151

Figure C.2 The reposition of LY10 to the docked fragments ........................................................................................................................................................................................................................................... 151

Figure C.3 The reposition of LY11 to the docked fragments ........................................................................................................................................................................................................................................... 152

Figure C.4 The reposition of LY17 to the docked fragments ........................................................................................................................................................................................................................................... 152

Figure D.1 Sesquiterpene lactone and its docking mode to STAT3 SH2 domain ........................................................................................................................................................................................................................................... 153

Figure D.2 Sesquiterpene lactone derivative and its docking mode to STAT3 SH2 domain ........................................................................................................................................................................................................................................... 154

Figure D.3 4-Beta-Hydroxywithanolide E and its docking mode to STAT3 SH2 domain ........................................................................................................................................................................................................................................... 154
List of Schemes

Scheme 7.1 Plausible mechanisms of the regioselective C-N coupling and C=N double bond ring fusion. ................................................................................................................................. 86

Scheme 7.2 Preparation of compound 1. Reaction Conditions: (a) NH$_3$·H$_2$O, Acetone, 0 °C to rt; (b) CrO$_3$, AcOH, 60 °C. ................................................................................................................................. 89

Scheme 7.3 Preparation of benzyl 4-(2-aminophenyl)piperazine-1-carboxylate. Reaction Conditions: TEA and DCM. ........................................................................................................... 90

Scheme 7.4 Control experiments on the reaction mechanism......................................................................................................................... 94

Scheme 7.5 Preparation of ring-opening 6-amino/anilino-naphthalene-dione-1-sulfonamide. Reaction Conditions: TFA and EtOH. .......................................................... 104
Chapter 1: Computer aided drug design

1.1 Introduction

Drug discovery is a highly risky and extremely time- and resource-consuming research area. Outcome is not positively proportional to the investment. Typically, it took about 14 years average to finish the cycle from concepts to market, and the cost is also approximately 1.0 billion dollar per drug\(^2\). Since the early 1990s, several technological and scientific advances, such as combinatorial chemistry, high-throughput screening, have been heralded as remedies to the problems facing the pharmaceutical industry. The use of these technologies in some form is now well established at most pharmaceutical companies; however, the return on investment in terms of marketed products didn’t meet expectations\(^3\). Because of the low cost and design-intensive features, computer aided drug design (CADD) is spread out as an alternative method to improve the low efficiency and high failure rate in drug discovery. CADD is a general term which summarizes all the \textit{in silico} methods applied in the drug discovery process, including molecular docking, high throughput virtual screening, combinatorial library design, \textit{in silico} fragment-based drug design (FBDD), molecular dynamic Simulation (MD), and induced-fit docking, etc.
1.2 Molecular docking and high throughput virtual screening

Molecular docking is a computational simulation which tries to predict the binding modes (position, orientation and conformation) between two or more substrates, including protein-ligand(s), protein-protein interactions. Scoring functions are used to numerically evaluate the preferred modes and analyze the strength of association or binding affinity between two/more molecules\textsuperscript{4}. For the target based drug discovery, molecular docking can assist to predict the interactions between biologically relevant molecules such as proteins, nucleic acids, lipids, and carbohydrates, etc. Thus docking plays a vital role in the rational drug design.

A typical docking experiment can be imagined as the problem of “lock-and-key”. For example, in Figure 1.1, the protein is the lock and the ligand is the key.

![Figure 1.1](image)

Figure 1.1 The “lock-and-key” diagram represents a typical docking experiment.
However, in human body, the biological processes are much more complicated than the relationship between “lock-and-key”. After ligand binding to protein, many possible situations might happen, for instance, 1. Ligand binding induces conformational change of protein, which enhances the binding affinity between the ligand and the protein. The functions of the protein are blocked. 2. After binding, protein like enzyme reacts with ligand which is either catalyzed by the enzyme or covalently bonded to the protein. 3. Ligand binding results in the conformational change of the allosteric binding sites of the protein. To deal with different situations, distinct computational methods are applied. For the above mentioned situations, a rigid docking experiment is not enough to solve the problems. For situation 1, flexible docking or induced-fit docking can be used. For situation 3, both molecular docking and dynamic simulations are required to solve the problems. For situation 2, molecular docking and dynamic simulations are not adequate to explain the biological effects. Further investigations such as structural analysis, chemical intuitions and other bench works are required to account for the phenomena.

Some widely used docking programs include AutoDock, Glide, Dock, GOLD, FlexX, etc. In this study, AutoDock and Glide are two primarily used docking softwares. And their scoring functions can be summarized as three types: force field based, empirical and knowledge based scoring functions.

High-throughput screening (HTS) approach can rapidly identify active compounds or lead molecules from millions of chemical, genetic or pharmacological tests. HTS is
popularly exploited in drug discovery, especially in pharmaceutical industry. Compared to HTS, high throughput virtual screening (vHTS) is a much faster, but less expensive and time consuming computational method\(^6\). vHTS is typically used at an early stage of the drug discovery process, including two approaches ligand and structure-based vHTS. When the structure of the target is unknown, ligand-based vHTS can be applied to search for active compounds based on the pharmacophore model built from existing known ligands. This step can also be set as a filter before a structure-based vHTS. Structure-based vHTS is a more straight-forward approach which can be considered as a pile of individual rigid docking experiments. Compounds from specific database are docked into the receptor binding site. Results are ranked by docking scores. Hits are selected according to docking scores and modes for further tests. Docking programs such as AutoDock and Glide can also be utilized in vHTS. Since Glide requires less computing time, in this study, Glide was used for the initial virtual screening. Selected hit library was cross-docked by Autodock. The false positive ligands were eliminated after the comparison of both results.
1.3 Fragment-based drug design (FBDD) and \textit{in silico} FBDD

FBDD, a drug design technique, under a number of different names (e.g. needles, shapes, binding elements, or seed templates), has actually been around for longer than immediate reference searches might suggest\textsuperscript{3}. For example, Card et al. introduced a fragment-based discovery method, but they referred as scaffold-based drug design\textsuperscript{7}. The timeline of fragment-based discovery can be generally summarized as following, 1. Scientists at Abbott used NMR to identify fragments binding to many protein drug targets, called SAR by NMR approach in the year 1996\textsuperscript{8}. 2. The first fragment-based drug discovery by X-ray crystallography was from another Abbott group in the year 2000\textsuperscript{9}. 3. FBDD was subsequently developed by various small technology companies for more focused application in drug discovery.

The general procedures of FBDD and vHTS were summarized in the following flowchart (Figure 1.2)\textsuperscript{10}. Minor revision might be needed for various cases. The screening of lower molecular weight and less complex fragments can result in lead compounds occupying more chemical spaces with better ligand efficiency, which accounts for the explosion in use of FBDD approaches in recent years.
Contrarily to HTS, where complete molecules are screened for activity, FBDD builds new ligands piece-by-piece by linking well-chosen fragments which bind to specific binding sites. FBDD is superior to HTS with higher chemical diversity, higher ligand efficiency, and higher hit rates. FBDD is also less resource-intensive than HTS. Acceptable IC<sub>50</sub> for fragments is < 1 mM; while acceptable IC<sub>50</sub> for lead-like molecules is < 1 μM. The comparison between these two methods is listed in Table 1.1.

In addition, FBDD offers several advantages over conventional diversity-oriented synthesis which greatly increases the efficiency of drug discovery and has more diversity<sup>11</sup>. Several compounds from FBDD are already in clinical trials (Table 1.2)<sup>12</sup>.
Table 1.1 Comparisons of HTS and FBDD.

<table>
<thead>
<tr>
<th><strong>FBDD</strong></th>
<th><strong>HTS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen a few hundred or thousand compounds</td>
<td>Screen hundreds of thousands of compounds</td>
</tr>
<tr>
<td>MW 150-300</td>
<td>MW 250-600</td>
</tr>
<tr>
<td>Biophysical screening techniques</td>
<td><em>In vitro</em> bioassay-based screening</td>
</tr>
<tr>
<td>Design-intensive</td>
<td>Resource-intensive</td>
</tr>
<tr>
<td>Find chemotypes that bind to difficult targets</td>
<td>not addressable</td>
</tr>
<tr>
<td>Hit rate around 1%(^{13})</td>
<td>Hit rate is 0.01% or less(^{13})</td>
</tr>
</tbody>
</table>
Table 1.2 Potent inhibitors derived from FBDD

<table>
<thead>
<tr>
<th>Company</th>
<th>Target</th>
<th>Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbott</td>
<td>Matrix metalloproteinase (MMP)</td>
<td>Phase I (ABT-516)</td>
</tr>
<tr>
<td>Abbott</td>
<td>B-cell CLL/lymphoma 2 (BCL-2), BCL-2-like 1 (BCL-XL)</td>
<td>Preclinical development (ABT-737)</td>
</tr>
<tr>
<td>Abbott</td>
<td>FK506-binding protein (FKBP)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Abbott</td>
<td>Leukocyte function-associated antigen-1 (LFA1)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Abbott</td>
<td>Protein tyrosine phosphatase-1B (PTP1B)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Abbott</td>
<td>Dihydrolipoyl dehydrogenase (DHNA)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Abbott</td>
<td>BCL-2 selective</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Abbott</td>
<td>Heat shock protein-90 (HSP90)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Abbott</td>
<td>Survivin</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Abbott</td>
<td>Poly (ADP-ribose) polymerase (PARP)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Abbott</td>
<td>Methionine aminopeptidase-2 (MetAP2)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Abbott</td>
<td>Casein kinase-2 (CK2)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Abbott</td>
<td>Kinase insert domain receptor (KDR)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Abbott</td>
<td>v-akt murine thymoma viral oncogene homolog 1 (AKT-1)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Astex Therapeutics</td>
<td>Aurora kinase[2]</td>
<td>Phase I (AT9283)</td>
</tr>
<tr>
<td>Astex Therapeutics</td>
<td>Cyclin-dependent kinase (CDK)</td>
<td>Phase I (AT7510)</td>
</tr>
<tr>
<td>Astex Therapeutics</td>
<td>CDK[4]</td>
<td>Preclinical development (AT9311)</td>
</tr>
<tr>
<td>Astex Therapeutics</td>
<td>HSP90[2]</td>
<td>Preclinical development (AT13387)</td>
</tr>
<tr>
<td>Astex Therapeutics</td>
<td>Mitogen-activated protein kinase-14 (P38c)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Aventis</td>
<td>Src SH2 domain[5]</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Novartis</td>
<td>3α-hydroxysteroid dehydrogenase (3α-HSD)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Plexikon</td>
<td>Peroxisome proliferator-activated receptor (PPARγ) (Metabolic)</td>
<td>Phase II (PPLM04)</td>
</tr>
<tr>
<td>Plexikon</td>
<td>Oncogenic v-raf murine sarcoma viral oncogene homolog B1 (B-Raf)</td>
<td>Investigational new drug application (PLX4032)</td>
</tr>
<tr>
<td>Plexikon</td>
<td>PPAR multiple sclerosis</td>
<td>Preclinical development</td>
</tr>
<tr>
<td>Plexikon</td>
<td>FMS/Kit</td>
<td>Preclinical development</td>
</tr>
<tr>
<td>Roche</td>
<td>DNA Gyrase</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>SGX Pharmaceuticals</td>
<td>SYK[8]</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>SGX Pharmaceuticals</td>
<td>Aurora kinase[9]</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Schering-Plough</td>
<td>β-site APP-cleaving enzyme-1 (BACE1)</td>
<td>Preclinical development</td>
</tr>
<tr>
<td>Schering-Plough</td>
<td>Mouse double minute-2 (MDM-2)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Schering-Plough</td>
<td>AKT-1[10]</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Schering-Plough</td>
<td>Hepatitis C virus (HCV) polymerase</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Sunesis</td>
<td>Interleukin-2 (IL2)[11]</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Sunesis</td>
<td>Caspase-3[15]</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Triad</td>
<td>P38c[17]</td>
<td>Preclinical development</td>
</tr>
<tr>
<td>Triad</td>
<td>c-Jun N-terminal kinase-2 (JNK2)[17]</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Vernalis</td>
<td>HSP90[18]</td>
<td>Preclinical development</td>
</tr>
<tr>
<td>Vernalis</td>
<td>Phosphoinositide-dependent protein kinase-1 (PDK1)</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Vernalis</td>
<td>Checkpoint kinase-1 (CK1)[19]</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Vernalis</td>
<td>Aurora kinase[20]</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Vertex</td>
<td>Jnk kinase-3 (JNK3)[21]</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Vertex</td>
<td>Adipocyte lipid-binding protein-2 (aF2)[22]</td>
<td>Novel, potent inhibitors</td>
</tr>
<tr>
<td>Vertex</td>
<td>Regulatory erythroid kinase (REDK)</td>
<td>Novel, potent inhibitors</td>
</tr>
</tbody>
</table>
To perform a successful FBDD, three key elements of FBDD need to be well-designed, including library of suitable fragments, methods to identify the binding affinity of fragments and strategy for evolving the fragments to larger hits for further optimization to lead compounds.

First, to build a qualified fragment library, molecules are selected based on “Rule of 3”, 1. MW < 300; 2. Number of hydrogen-bond donors ≤ 3; 3. Number of hydrogen-bond acceptors ≤ 3; 4. cLogP ≤ 3; 5. The number of rotatable bonds was, on average, ≤ 3; 5. Polar surface area was ≤ 60 Å². Other criteria are also used to build a fragment library, such as water solubility, maximization of the diversity of a given library. To enrich the diversity, methods like the scaffold-based classification approach (SCA) can be used

Second, several screening methods are available to identify the binding affinity of fragments, listed in Table 1.3.

To further increase the hit rates, combination of two to three screening methods can improve the hit rate to about 2-10%, such as biochemical screening and NMR; NMR and X-ray; SPR and X-ray; NMR, SPR and X-ray, etc.
Table 1.3 The comparison of the screening methods.

<table>
<thead>
<tr>
<th><strong>Methods</strong></th>
<th><strong>Advantages</strong></th>
<th><strong>Disadvantages</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray crystallography</td>
<td>Supply immediate structural information</td>
<td>Satisfy both the weak binding and crystal formation, library size &lt; 1000</td>
</tr>
<tr>
<td>NMR</td>
<td>Ligand-protein interactions</td>
<td>library size &lt; 2000</td>
</tr>
<tr>
<td>SPR (Surface Plasmon Resonance)</td>
<td>Sensitivity and low material consumption, label-free</td>
<td>Tethering of molecules to immobilize the protein may affect the binding constants measured.</td>
</tr>
<tr>
<td>FCS (Fluorescence correlation spectroscopy)</td>
<td>Easy to be detected, able to detect single molecule, library size &gt; 10,000</td>
<td>Labeling ligands or protein is cost intensive and may interfere with the binding</td>
</tr>
<tr>
<td>Mass-spectrometry</td>
<td>rapid analysis, high sensitivity, low sample consumption</td>
<td>hampered by low molecular weight matrix derived interference signals</td>
</tr>
<tr>
<td>ITC (Isothermal titration calorimetry)</td>
<td>Used to detect enthalpy change in binding</td>
<td>Requires a lot of protein</td>
</tr>
</tbody>
</table>
Table 1.3 continued

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity and low material consumption</th>
<th>The fluorescence signal from the substrates causes the deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP (Fluorescence Polarization)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MST (Microscale thermophoresis)</td>
<td></td>
<td>High cost</td>
</tr>
<tr>
<td><em>in silico</em> screening</td>
<td>Library size &gt; 1,000,000, No sample consumption</td>
<td>Hit rate is comparatively low</td>
</tr>
</tbody>
</table>

Third, evolving the fragments to larger hits for FBDD uses three general strategies, growing, linking and merging. As shown in Figure 1.2\textsuperscript{15}, growing and linking are based on selected fragments; merging needs one additional step to cut the existing inhibitors into fragments for further recombination.

![Figure 1.3 General strategies for evolving the fragments to larger hits](image)

Figure 1.3 General strategies for evolving the fragments to larger hits\textsuperscript{9}.  

11
In silico FBDD utilizes the computational methods to perform the virtual FBDD. To improve the efficiency of in silico FBDD, scaffolds found in natural products, amino acids and approved drugs have higher priority to be selected into the fragment library for the FBDD library design. Many available computational programs can assist us to do further fragment screening and evolving procedures, such as, multiple ligands simultaneous docking (MLSD), Combiglide, Allegrow, Ludi, LigBuilder, etc. For the fragment selection, programs like Combiglide and Allegrow, can be used to search for suitable fragments for further growing and merging step; and MLSD can be used to simultaneously search for multiple fragments for the linking method. Combiglide and MLSD programs were used in our fragment-based drug discovery study.

1.4 CombiGlide and Multiple Ligand Simultaneous Docking (MLSD)

In most cases, de novo drug discovery provides lead molecules, which need further optimization to become drugs in market. Schrödinger's CombiGlide creates virtual combinatorial libraries that generate focused libraries in support of lead optimization efforts. CombiGlide is a combinatorial technology which operates core hopping for lead discovery and optimization. According to the analysis of the protein-ligand interaction, by replacing existing groups or by attaching new fragments to core structures, the fragment with weak binding affinity is replaced and the binding affinity of the original
molecule is enhanced by adding a fragment with high ligand efficiency. CombiGlide (Schrodinger LLC)\textsuperscript{16} was applied to design focused libraries and to modify the known STAT3 inhibitor XZH-5. “Fragment-like" subsetID #2 from Zinc Libraries (701,337 entries) was used for reagent preparation.

MLSD as an optimized FBDD computational program was developed by our group\textsuperscript{17} which simulates multiple (more than two) molecules’ recognition processes by improved sampling method and scoring function. In this thesis, MLSD was used to search for new STAT3 inhibitors for anticancer therapy.
1.5 Molecular Dynamics Simulation (MD)

Molecular dynamics (MD) simulates physical movements of atoms and molecules which are applied in the modeling of biomolecules. The trajectories of atoms and molecules are most commonly determined based on the Newton's equations of motion for a system of interacting particles. Biomolecules are wrapped by explicit solvent molecules in the MD simulation, which are very valuable for understanding the dynamic behavior of solvent influence. MD can not only predict movements from fast internal motions to slow conformational changes or even protein folding processes, but also possible to study the stability of the biomolecular system, such as density, conductivity, and dipolar moment, as well as different thermodynamic parameters, including interaction energies and entropies\(^{18-20}\).

In this thesis, Amber program is used to do MD simulation of STAT3 protein and its inhibitors. By the analysis of MD simulation, we can find out the induced conformations of STAT3 and perform further free energy calculations.
1.6 Induced-fit docking (IFD)

In standard virtual docking studies, flexible moving ligands are docked into the binding site of a rigid receptor, which can be a crystalized structure or a homologized structure with lowest energy. However, in reality, the energy of binding receptor may not be the lowest state as the crystalized structure. In addition, the binding process involves various degrees of movements of both ligand and receptor. The receptor alters its binding site to fit the ligand better and to lower the energy of the complex, so that this process is referred as “induced fit”\(^1\). Modeling induced-fit docking can generate an accurate docking mode of a known ligand but that cannot be obtained by the docking experiment on a rigid structure of the receptor. IFD can both eliminate the false positives and rescue the false negatives in virtual screening experiments. The accuracy of current available programs for IFD is still not very satisfying. The use of active ligand sets and decoy ligand sets of a specific receptor is a more reliable approach to test the induced structure.
1.7 Free energy calculation

To identify the free binding energy contribution of the residues in hot spots of receptor, MM-GB/PBSA approach can be applied to search for key residues which contribute to the majority free energy of ligand-receptor binding. Ideally the binding energy ($\Delta G_{binding}$) is the free energy difference between the bound and unbound states of two solvated molecules calculated by equation 1.

$$[\text{ligand}]_{sol} + [\text{receptor}]_{sol} \rightarrow \text{plex}_{sol}$$  \hspace{1cm} (1)

However, since the majority of the energy contribution comes from water-water interactions so that the fluctuations in total energy will be an order of magnitude larger than binding energy, in MM-GB/PBSA, the binding energy of ligand-receptor complex is calculated according to the following thermodynamic cycle by equations (2-3):

$$[L]_{sol} + [R]_{sol} \xrightarrow{\Delta G_{binding,sol}} \chi[L - R]_{sol}$$

$$\uparrow \Delta G_{L,sol} \hspace{1cm} \uparrow \Delta G_{R,sol} \hspace{1cm} \uparrow \Delta G_{L-R,sol}$$ \hspace{1cm} (2)

$$[L]_{gas} + [R]_{gas} \xrightarrow{\Delta G_{binding,gas}} [L - R]_{gas}$$

$$\Delta G_{binding,sol} = \Delta G_{L-R,sol} - \Delta G_{L,sol} - \Delta G_{R,sol}$$ \hspace{1cm} (3)

The binding energy equals to the sum of the difference of enthalpy and entropy. Enthalpy ($\Delta H_{binding,sol}$) can be further decomposed into individual contributions from MM element ($E_{MM}$), GB/PB element ($G_{GB/PB}$) and SA element ($G_{SA}$). The MM element includes $E_{int}$ (bond, angle, and dihedral energies), $E_{ele}$ (electrostatic energy) and $E_{vdw}$ (Van der Waals).
is the sum of electrostatic solvation energy using different solvation models, implicit solvent model (GB) and explicit solvent model (PB). \(G_{SA}\) is the non-electrostatic solvation energy including the cavity formation and Van der Waals interactions between the solute and the solvent. It can be estimated using solvent-accessible surface area (SASA) with parameters, the surface tension proportionality constant, \(\gamma = 0.0072 \text{ kcal mol}^{-1} \text{ Å}^{-2}\), and the free energy of non-electrostatic solvation energy for a point solute, \(b = 0.00 \text{ kcal mol}^{-1}\). Equations (4-7) are summarized as below:

\[
\Delta G_{\text{binding,sol}} = \Delta H_{\text{binding,sol}} - T \Delta S_{\text{binding,sol}} \quad (4)
\]

\[
G = E_{\text{MM}} + G_{\text{GB/PB}} + G_{SA} - TS \quad (5)
\]

\[
E_{\text{MM}} = E_{\text{int}} + E_{\text{ele}} + E_{\text{vdw}} \quad (6)
\]

\[
G_{SA} = \gamma \times \text{SASA} + b \quad (7)
\]

To identify the residues of hot spots which contribute most of the binding energy, pairwise residue free energy decomposition of MMGBSA can be performed. Then the decomposed binding free energy of all the amino acid residues can be sorted. Finally, the residues which contribute the most of the negative free energy can be selected as hot spot residues\(^{22-25}\).
Chapter 2: Biochemical assays

2.1 Introduction

For CADD, a big challenge is how to validate the calculated results. Cell based biological assays reflect complicated biological processes in vitro, which are indirectly correlated to the binding effects between drug and target. So the derived SAR or QSAR based on biological assays usually suffer with low accuracy. To better modify existing active compounds, binding constants of active compounds can be used as a critical criterion, which can be measured by target based biochemical assays. Biochemical assays can either directly or indirectly test the binding affinity between target and ligand, including various methods, Fluorescence Polarization (FP) assay, Microscale thermophoresis (MST) assay, Surface Plasmon Resonance (SPR), Isothermal Titration Calorimetry (ITC), etc. In our work, we used FP and MST to measure the binding constants of our inhibitors to STAT3.

2.2 Fluorescence Polarization (FP) assay

The technique of fluorescence polarization (FP) is based on the observation that when a fluorescently labeled molecule is excited by polarized light, it emits light with a degree of polarization that is inversely proportional to the rate of molecular rotation. This property of fluorescence can be used to measure the interaction of a small labeled ligand with a larger protein and provides a basis for direct and competition binding assays.\textsuperscript{26,27}
A fluorophore is covalently attached to the native N-terminus peptide, for example, 5-carboxyfluorescein-GY(PO3H2)LPQTV-NH2. It is excited by polarized light; the emitted light will be largely depolarized. If this labeled peptide is bound to STAT3 protein, the fluorophore reorients to a much smaller degree, due to the significantly reduced rotational speed of the complex. Thus, the emitted light will still be polarized to a significant degree.

After adding the designed inhibitor to the system, it can compete with the labeled peptide to bind to target protein, so that the alternation of FP reading values is correspondent to the inhibition percentage of the inhibitor.
2.3 Microscale thermophoresis (MST) assay

Microscale thermophoresis (MST) is a technology based on a physical principle used in bioanalytics. It detects the changes of biomolecular hydration shell due to structural change resulting in a correlative change of movement along the temperature gradient. This theory is used to determine binding affinities, binding kinetics and activity kinetics. It allows fast detection of a wide range of biomolecular interactions\textsuperscript{28,29}.

The main feature of MST is that measurements can be performed label-free, immobilization-free and in any buffer. However, the cost of MST test is very high. One sample costs at least $100 only for the reagents and facilities, not to mention the cost of protein and inhibitors, which limits the spread of MST method. In this thesis, MST method was used to measure the binding affinity between STAT3 and the designed inhibitors. Several new compounds were discovered by MST, represented by inhibitor LY17.
Chapter 3: Brief background of STAT3 and its signaling pathways

The family of signal transducers and activators of transcription (STAT) consists of seven proteins in humans (STAT1-4, STAT5A, STAT5B, and STAT6). STAT2, STAT4, STAT6 are related to the development of T-cells and in IFN-gamma signaling pathway. STAT1, STAT3, STAT5A, STAT5B are most related to human cancer progression. STAT knock-out mice are viable and fertile, except for STAT3 knock-out mice, which are embryonic lethal early in development.

Through both intrinsic and extrinsic pathways, STAT3 is involved in the development of inflammation and cancer. The intrinsic pathway includes the overexpression of growth factor receptors with intrinsic tyrosine kinase activity and cytokine receptors with associated Janus kinase (JAK) family tyrosine kinases. The extrinsic pathways includes oncogenic mutations in JAK family members, activated non-receptor tyrosine kinases such as SRC, and induced membrane associated receptors by environmental factors, including ultraviolet (UV) radiation, chemical carcinogens, infection, stress and cigarette smoke. Activated STAT3 forms dimers that translocate to the nucleus, bind to DNA and regulate gene expression. STAT3 induces the expression of many cytokines, chemokines and other mediators, such as interleukin-6 and cyclooxygenase 2, while chemokines and mediators in turn further activate STAT3, forming autocrine and paracrine feedforward loops, as shown in Figure 3.1.
Figure 3.1 The activation of STAT3 signaling pathway by intrinsic and extrinsic pathways\textsuperscript{30}. 

Signal Transducers and Activators of Transcription 3 (STAT3) is an anti-cancer target, which is constitutively activated in many kinds of human tumor systems. Phosphorylated Tyr-705 (pTyr705) of both STAT3 monomers can bind to each other’s Src Homology 2
(SH2) domain to form a homo-dimer. The STAT3 dimer can bind to DNA chain, then activate cell proliferation and finally develop into cancer \(^{32}\) (Figure 3.2A).

Figure 3.2A. The Tyr705 residues of STAT3 monomers are phosphorylated then the homo-dimer binds to DNA, and finally cause cancer. B. The chemical structure of XZH-5. C. The binding mode of XZH-5 predicted by Autodock4.

Our objective of this study is designing the promising homo-dimerization inhibitors for STAT3 protein, assisted by computationally generated optimal protein binding conformation, evaluated by tested active compounds.
Chapter 4: Discovery of novel STAT3 small molecule inhibitors via in silico site-directed fragment-based drug design


4.1 Abstract

Constitutive activation of Signal Transducer and Activator of Transcription 3 (STAT3) has been validated as an attractive therapeutic target for cancer therapy. To block both STAT3 activation and dimerization, a viable strategy is to design inhibitors competing with the native phosphotyrosine peptide that binds to the STAT3 SH2 domain.

An improved fragment-based drug design (FBDD) strategy, in silico site-directed FBDD, was applied in this study. The binding between 5,8-dioxo-6-(pyridin-3-ylamino)-5,8-dihydronaphthalene-1-sulfonamide (LY5) and STAT3 SH2 domain was confirmed by fluorescence polarization assay. Four out of the five selected and synthesized compounds have IC\textsubscript{50} values lower than 5 µM for the U2OS cancer cells. 8 (LY5) has an IC\textsubscript{50} range
in 0.5-1.4 µM in various cancer cell lines. 8 also suppresses tumor growth in vivo. This study has demonstrated the feasibility of in silico site-directed FBDD for STAT3 drug discovery; and can be used to other drug targets in general.

4.2 Introduction

Constitutive activation of STAT3 has been found in a wide variety of cancers, including breast cancer, sarcomas, and other cancers, promoting it as a very attractive therapeutic target. Cytokines, hormones, and growth factors binding to the cell surface receptors can activate the JAK-STAT signaling pathway. Activated receptors activate JAK kinase(s) and autophosphorylate themselves. Subsequently, the STAT3 monomer is phosphorylated at Tyrosine705 (pTyr705) by the same kinases through STAT3 SH2 domain binding to pTyr loop of the activated receptors, leading to STAT3 homodimer through its SH2 dimerization. The dimerized STAT3 then translocates into the nucleus and binds to DNA, turning on a host of oncogenes. Altogether, these events lead to cell proliferation, apoptosis resistance, etc. To block both phosphorylation and dimerization processes, STAT3 inhibitors should compete with the native phosphotyrosine (pTyr705) peptide by binding to the STAT3 SH2 domain (Figure 4.1).
Figure 4.1 Docking model of STAT3 inhibitor 8 binding to the STAT3 SH2 domain (PDB: 1BG1), generated by AutoDock4 and viewed by Maestro. The two key binding pockets are called “site pTyr705” (red circle) and “side pocket” (blue circle). Carbon atoms of 8 are colored purple, and those of the pTyr peptide are colored green. The style of molecular surface is mesh, colored according to element property.

Several series of STAT3 dimerization inhibitors have been discovered via both computational and experimental methods. 1) Phosphopeptide mimics were initially developed as STAT3 inhibitors to compete with the native phosphopeptide of the STAT3 protein. For example, PM-73G is a phosphopeptide mimic STAT3 inhibitor that can completely inhibit STAT3 Tyr705 phosphorylation at 0.5-1µM level in various cancer cell lines. Another phosphopeptide mimic, pCinn-Leu-cis-3,4-methanoPro-Gln-NHBn, has the lowest reported IC$_{50}$ value at 69 nM, as determined by fluorescence polarization.
2) Peptidomimetics were also designed to target the STAT3 SH2 domain. Peptidomimetics are derived from phosphopeptides that mimic peptides but do not necessarily contain phosphate groups. For example, XZH-5 was designed using a structure-based approach to inhibit the formation of STAT3 dimers\textsuperscript{35,36}. 3) Various small molecules have been reported to inhibit STAT3 dimerization, making them as more druggable candidates. STA-21 discovered by structure-based virtual screening was the first reported small inhibitor. It inhibits STAT3 dimerization, DNA binding, and STAT3-dependent luciferase reporter activity in breast cancer cells\textsuperscript{37}. Another small molecule, Stattic, was discovered by high-throughput screening and has been shown to selectively inhibit activation, dimerization, nuclear translocation of STAT3, and to increase apoptosis in STAT3-dependent cancer cell lines\textsuperscript{38}. Among all the reported non-peptidomimetic small inhibitors, 5-hydroxy-9,10-dioxo-9,10-dihydroanthracene-1-sulfonamide (LLL12) has the lowest \textit{IC}_50 (0.16-3.09 µM\textsuperscript{39}), inhibiting STAT3 phosphorylation and the growth of human cancer cells. However, so far, there is no STAT3-targeting drug approved by the FDA. The search for more druggable STAT3 inhibitors with higher potency and better bioavailability remains extremely important.

Fragment-based drug design (FBDD) has emerged as a new strategy for drug discovery in the past decade. FBDD as a design-intensive method is superior to the resource-intensive conventional drug discovery methods, such as high-throughput screening and combinatorial chemistry\textsuperscript{40}. Conventional FBDD has made some improvement on efficiency and cost-effectiveness in drug design and several drug candidates designed by this methodology are currently under clinical trials. Computational fragment-based drug
design is another vibrant field. Chen and Shoichet utilized the computational fragment-based approach and successfully discovered potent *de novo* inhibitors against CTX-M, while the docking screens of lead-like compound libraries failed to identify any lead-like inhibitors.

In this article, we present the concept of *in silico* site-directed FBDD as a computational FBDD approach too, summarized in Figure 4.2.

Figure 4.2 Site-directed FBDD strategy. Step 1: Fragment libraries are categorized from existing inhibitors according to their binding modes. Step 2: New lead library is merged from randomly selected fragments from site-specific fragment sub-libraries. (Any existing inhibitors are removed from the new lead library.) Leads are selected after
evaluating docking modes to the target. Step 3: Final hits are identified after synthesis and testing.

With our *in silico* approach, fragment libraries are built from known inhibitors and are divided into binding site specific sub-libraries according to docking poses. Linkers from the known inhibitors are the first choices to join fragments, however new linkers based on the concept of bioisosterism\textsuperscript{42} may also be applied to join fragments. New linkers designed in this fashion can often maintain the original binding interactions or even enhance binding. Merging, the joining of fragments and linkers is different in our computational approach compared to conventional FBDD. Conventional approaches randomly merge fragments together while our pre-sorted, site-specific fragment sub-libraries are recombined to maximize the possibility of obtaining high affinity lead molecules. Merged potential candidates may be quickly screened via computational docking methods to further narrow the number of molecules that are synthesized and tested. The comparisons between conventional and *in silico* site-directed FBDD methods are summarized in Table 4.1.
### Table 4.1 The comparison of conventional FBDD and in silico site-directed FBDD

<table>
<thead>
<tr>
<th></th>
<th>Conventional FBDD</th>
<th><em>In silico</em> site-directed FBDD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fragments library</strong></td>
<td>Screened by X-ray, NMR and other methods</td>
<td>Built from known inhibitors and divided into binding site specific sub-libraries according to docking poses</td>
</tr>
<tr>
<td><strong>Linkers</strong></td>
<td>Selected from linker libraries or from chemical intuition</td>
<td>Original linkers or bioisosteres derived from original linkers</td>
</tr>
<tr>
<td><strong>Merge methods</strong></td>
<td>Fragments are randomly merged together</td>
<td>Fragments selected from sub-libraries are merged together</td>
</tr>
<tr>
<td><strong>Pre-selected candidates</strong></td>
<td>According to structural variety, purchase/synthesis availability or chemical intuition</td>
<td>According to computational docking results</td>
</tr>
</tbody>
</table>

To test the method, we have applied the approach to the case of STAT3 inhibitor design. The STAT3 fragment library was generated from nine known inhibitors with proven affinity and established pharmacological activities for STAT3. New leads were designed to target site 705 and the side pocket (Figure 4.1). The STAT3 fragment library was divided into two specific sub-libraries (site 705 and side pocket) based on the docking poses of the inhibitors to the crystal structure of STAT3 SH2 domain (PDB: 1BG1).

To efficiently evolve the fragments into leads, several considerations were made when choosing the linker and performing the merging. A desirable linker should allow sufficient flexibility for the fragments to maintain their poses in the binding sites and
enhance binding affinity and/or biophysical features such as water solubility. Most importantly, the chosen linker should not complicate synthesis. In this case, a secondary amine was chosen as the linker. The merged candidates were then screened via computational docking, and the compounds with the most favorable docking energies and well-clustered binding poses were selected for synthesis and experimental testing. This led to a new class of STAT3 inhibitors.

4.3 Results

4.3.1 In silico site-directed drug design

Step 1: Fragment libraries were categorized from the known STAT3 dimerization inhibitors based on their binding modes.

To build fragment libraries from the known STAT3 inhibitors\textsuperscript{35,37-39,43-47}, the compounds were docked to the STAT3 SH2 domain (PDB: 1BG1). The initial docking results are shown in Figure 4.3. Based on the binding modes, the fragments were divided into libraries specific for each of the two binding sites: site pTyr705 and side pocket (Figures 4.4 and 4.5).
Figure 4.3 Docking modes of the selected known STAT3 inhibitors with the STAT3 SH2 domain (PDB: 1BG1), generated by AutoDock4. Surface representation was created with Maestro.
Figure 4.4 Chemical structures of the selected STAT3 dimerization inhibitors. According to the docking modes in Figure 4.3, the fragments binding to the pTyr705 site are colored red, and those binding to the side pocket are colored blue.
Figure 4.5 Categorized fragment sub-libraries, site pTyr705 (1-7) and side pocket (a-g).
(The fragment phosphotyrosine group was not included in the fragment libraries due to its peptidomimetic property.)

**Step 2: New lead library was built by linking selected fragments from different fragment sub-libraries.**

Based on the docking modes for known STAT3 inhibitors, naphthalene-5,8-dione-1-sulfonamide (1) exhibited the most interactions with site pTyr705 by forming two hydrogen bonds with Arg609 and an additional hydrogen bond with Ser613. Additionally, as part of 13 (LLL12), one of the most potent small nonpeptidomimetic inhibitors of STAT3 dimerization, 1 is an attractive starting point for the synthesis of new
STAT3 inhibitors. Therefore, 1 was selected as the component targeting site pTyr705 and was randomly linked to fragments from the side pocket sub-library to form new potential inhibitors.

The linker was designed based on the rationales described previously, that the new linker can maintain or even enhance the binding affinity of the two fragments and that the synthetic strategy would be feasible. Based on the docking mode of 13 in Figure 4.3, the phenol ring of LLL12 is the original linker of 1. Given the difficulty of synthesis, the original linker was simplified into an isopropyl group which was subsequently evolved into an isopropyl amine group based on bioisosterism. To further reduce the synthetic difficulty, the isopropyl amine was modified to a dimethyl amine linker as shown in Figure 4.6. Finally, the dimethyl amine group was selected as a suitable linker after docking results indicated that candidates with this linker can form hydrogen bonds with the backbone of Ser636 in a similar fashion as the hydroxyl group of 13 (Figure 4.7).

Figure 4.6 The evolution of linker design.
Figure 4.7 The docking modes of 13 (LLL12) and 8 (LY5). Both linkers, the hydroxyl group in 13 and secondary amine group in 8, form hydrogen bonds with the Ser636.

**Step 3: Final leads for further synthesis and tests were selected by repositioning the compounds from lead library to STAT3 SH2 domain.**

Docking the new lead library to the STAT3 SH2 domain, compounds were ranked based on the docking scores and clustering, and compounds that cannot reposition to the binding pockets were removed from consideration. The compounds that were ultimately selected for synthesis and testing are shown in Table 4.2.
Table 4.2 Chemical structures, docking scores, and IC$_{50}$ of the designed STAT3 inhibitors

<table>
<thead>
<tr>
<th>Cpd</th>
<th>R</th>
<th>Docking score$^a$ (kcal/mol)</th>
<th>IC$_{50}$$^b$ (µM)</th>
<th>Cpd</th>
<th>R</th>
<th>Docking score (kcal/mol)</th>
<th>IC$_{50}$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td></td>
<td>-7.3</td>
<td>0.5</td>
<td>10</td>
<td></td>
<td>-7.1</td>
<td>5.0</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>-7.2</td>
<td>&gt;5.0</td>
<td>11</td>
<td>Cl</td>
<td>-8.4</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td></td>
<td>-8.0</td>
<td>2.5</td>
</tr>
</tbody>
</table>

$^a$Docking scores were calculated by AutoDock4. $^b$IC$_{50}$ was tested on sarcoma cell line U2OS.

4.3.2 Biological and biochemical assays.

Cell Viability Assay. Initially, IC$_{50}$ values for all of the compounds were tested with U2OS cell line. Then two most potent compounds, 8 (LY5) and 11 were further tested together with 13 in two other RD2 and RH30 cell lines. As 13 had the lowest IC$_{50}$ among
the known small inhibitors, it was selected as a control compound. The results showed that 8 had lower IC$_{50}$ than 13 in cell lines RD2 and U2OS and similar IC$_{50}$ to 13 in cell line RH30. 11 had a comparable IC$_{50}$ as 13 in cell lines RD2 and U2OS (Table 4.3).

Table 4.3  IC$_{50}$ values for human sarcoma cells (RD2, RH30 and U2OS) inhibition by STAT3 inhibitors, 13 (LLL12), 8, and 11.

<table>
<thead>
<tr>
<th>Compd</th>
<th>U2OS</th>
<th>RH30</th>
<th>RD2</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>1.00</td>
<td>0.47</td>
<td>2.85</td>
</tr>
<tr>
<td>8</td>
<td>0.52</td>
<td>0.55</td>
<td>1.39</td>
</tr>
<tr>
<td>11</td>
<td>1.39</td>
<td>1.14</td>
<td>2.56</td>
</tr>
</tbody>
</table>

**Compound 8 Inhibits STAT3 Phosphorylation and Induces Apoptosis in Human Sarcoma Cancer Cells.** To examine the inhibition of STAT3 phosphorylation, Western blot assays were performed to detect the abundance of phosphorylated STAT3 (P-STAT3) after RH30 cells were treated with 11 (0.5-2.5 μM) or 8 (0.25-1 μM) for 16h. STAT3 phosphorylation was reduced in a dose-dependent manner, with 8 and 11 almost completely inhibiting Tyr705 phosphorylation at 0.5 μM and 1 μM, respectively (Figure 4.8). Since 8 was shown to be a more potent inhibitor of STAT3 phosphorylation than 11 in the RH30 cell line, it was further tested in EW8 sarcoma cells. Western blots of EW8 cells treated with 8 (0.5-1.0 μM) for 8h again reveals a dose-dependent decrease in
formation of P-STAT3 (Figure 4.9). The expression of total STAT3 was not changed in both RH30 and EW8 cell lines, indicating that the decrease of P-STAT3 was not due to a constitutional decrease of total STAT3 expression. 8 was not found to inhibit phosphorylation of the kinase ERK1/2. The inhibition of STAT3 phosphorylation by 8 seems to be consistent with the induction of apoptosis as evidenced by the presence of cleaved caspase 3 (Figure 4.9).

![Western blot assay](image)

Figure 4.8 11 and 8 inhibit constitutive STAT3 phosphorylation. Human Rhabdomyosarcoma cell line RH30 was treated with 11 (0.5-2.5 µM) and 8 (0.25-1 µM) for 16h, then whole-cell extracts were prepared and phospho-STAT3 was detected by Western blot assay, revealing a decrease in STAT3 phosphorylation.
Figure 4.9 Compound 8 inhibits constitutive STAT3 phosphorylation. Ewing’s sarcoma cell line EW8 was treated with 8 (0.5-1 µM) for 8h, then whole-cell extracts were prepared and phospho-STAT3 was detected by Western blot assay, revealing a decrease in STAT3 phosphorylation.

**Compound 8 Inhibits STAT3 Phosphorylation Induced by IL-6.** IL-6 can induce the activation of STAT3\(^{48}\). The MCF-7 breast cancer cell line was used to determine the ability of 8 to inhibit IL-6 induced STAT3 phosphorylation, since MCF-7 cells do not express persistently phosphorylated STAT3. In control experiments, IL-6 elevated STAT3 phosphorylation in MCF-7 cells, while 8 and 11 blocked the stimulation, with phosphorylated STAT3 abrogated at low concentrations (Figure 4.10). The DAOY medulloblastoma cancer cell line was used to further determine the ability of 8 to inhibit
IL-6 induced STAT3 phosphorylation. In control experiments, IL-6 elevated STAT3 phosphorylation in DAOY cells, while 8 blocked the stimulation, with phosphorylated STAT3 abrogated at low concentrations (Figure 4.11).

<table>
<thead>
<tr>
<th>IL-6 (25 ng/mL)</th>
<th>0</th>
<th>0.25</th>
<th>0.5</th>
<th>1.0</th>
<th>0.5</th>
<th>1</th>
<th>2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 (µM)</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>11 (µM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Figure 4.10 11 and 8 inhibit IL-6 induced STAT3 phosphorylation in MCF-7 breast cancer cells. The MCF-7 cells were serum-starved overnight, then left untreated or treated with 11 (0.5-2.5 µM) or 8 (0.25-1 µM) for 5h, followed by stimulation by IL-6 (25 ng/mL). The cells were harvested at 30 minutes and analyzed by Western blot assays.
Figure 4.11 8 inhibits IL-6 induced STAT3 phosphorylation in Daoy cancer cells. The Daoy cells were serum-starved overnight, then left untreated or treated with 8 (1-5 µM) for 2h, followed by stimulation by IL-6 (50 ng/mL). The cells were harvested at 30 minutes and analyzed by Western blot assays.

**Compound 8 Inhibits STAT1 Phosphorylation Induced by IFN-γ.** IFN-γ can induce the activation of STAT1. The selectivity of 8 for STAT3 inhibition was evident when compared with that for STAT1. IFN-γ treatment elevated STAT1 phosphorylation in DAOY cells, whereas 8 pre-treatment had no effects on the extent of STAT1 phosphorylation (Figure 4.12).
Figure 4.12 IFN-γ induced STAT1 phosphorylation in Daoy cancer cells. The Daoy cells were serum-starved overnight, then left untreated or treated with 8 (1-5 µM) for 2h, followed by stimulation by IFN-γ (50 ng/mL). The cells were harvested at 30 minutes and analyzed by Western blot assays.

**Compound 8 Suppresses Tumor Growth of Breast Cancer–initiating Cells in Mouse Tumor Model in vivo.** We performed in vivo studies to confirm the tumor suppressor activity of 8 and its potential usage for breast cancer treatment. MDA-MB-231 cells were injected subcutaneously into nude mice for the induction of xenografts. After a 21 days treatment, the morphological data showed that 8 suppressed xenograft growth and significantly (P < 0.001) shrank the tumor sizes (Figure 4.13).
Figure 4.13 Compound 8 suppresses tumor growth of MDA-MB-231 breast cancer-initiating cells in mouse tumor model in vivo. Tumor growth was determined by measuring length (L) and width (W) of the tumor every other day with a caliper. The tumor volume was calculated according to the formula: Tumor volume=0.5236×L×W². The treatment lasted for 21 days. The results showed that 8 significantly suppresses the tumor growth (P < 0.001).

Fluorescence Polarization (FP) Assay. FP assay was used to further prove the binding of 8 to the STAT3 SH2 domain. Screening was performed at physiologically relevant temperature 37°C. Before each experiment, 8 was prepared freshly from the stock solution. Compound 8 increased its background fluorescence reading at 595nm at high concentrations, so the control solutions for each concentration were prepared by mixing buffer, 8, and fluorescence-labeled peptides together. Inhibition curves were fitted in
SigmaPlot11.0 (Figure 4.14) and $K_i$ value was determined using a mathematical model developed for FP assay at 2.5 µM. So the experimental $\Delta G$ value of 8 is -7.9 kcal/mol, which is comparable to the AutoDock4 computed $\Delta G$ value, -7.3 kcal/mol.

Figure 4.14 Compound 8 binds to STAT3 SH2 protein. Inhibition of binding of fluorescein-labeled phosphopeptides to the SH2 domains of STAT3 by 8 at 37 °C was assayed by fluorescence polarization. Error bars represent standard deviation.

4.4 Discussion
In this work, we introduce the concept of *in silico* sited-directed FBDD that utilizes methods of fragment-based drug design, *in silico* docking, and bioisosterism to recombine fragments of known inhibitors to form new lead molecules. The successful application of this method is detailed in the design of a novel STAT3 inhibitor scaffold, leading to compounds with low IC$_{50}$ values and strong binding affinity to STAT3 SH2 domain. This study has several significant findings: 1) The use of fragment sub-libraries based on the docking modes of known inhibitors was an efficient method to design potential new inhibitors. This method is potentially more reliable than virtual screening of fragment libraries and less costly than conventional methods like high-throughput screening. It is also distinct from the molecular hybridization$^{30,49}$, as in our approach, fragments are linked together based on structural information of specific target and the linked new compound is designed to bind to the same target. In addition, the final docking step of the merged fragments acts as an extra filter to assure that key interactions are maintained or even optimized prior to synthesis. 2) The secondary amine linker used in the design is a bioisostere of the hydroxyl group in 13 that can maintain both the binding poses of the fragments and the binding affinity. 3) The synthesis of novel STAT3 inhibitors presents a new highly regioselective coupling method. The major product is connected at position 6 of the napthoquinone. Compared to 13, the novel series of STAT3 inhibitors are superior with lower synthetic cost and larger space for further modification. Compound 13 has some disadvantages such as the need for the expensive synthetic material 3-hydroxy-2-pyrone, time-consuming purification to separate the regioisomers, and difficulty in modifying the three ring system. The novel STAT3 inhibitors can
overcome these shortcomings. The synthetic materials for novel STAT3 inhibitors are about 10 times less costly, the reaction has much higher selectivity to react to 6-naphthalene-5,8-dione-1-sulfonamide than to 7-naphthalene-5,8-dione-1-sulfonamide, and it is very easy to modify the novel STAT3 inhibitors to fit the side pocket by alternating the amine reagents. 4) For protein level study, the fluorescence polarization (FP) assay proved that 8 binds to STAT3 SH2 domain with a $K_i$ value of 2.5µM. 5) For 

*in vitro* study, the IC$_{50}$ of 8 is superior to that of 13 in two different cancer cell lines, U2OS and RD2, while 13 has the lowest IC$_{50}$ among all the reported nonpeptidomimetic small molecule inhibitors targeting SH2 domain. 6) For DAOY cell line, 8 reveals a clear selectivity between IL-6 induced P-STAT3 and IFN-γ induced P-STAT1, shown in Figure 4.11 and 4.12. 7) For *in vivo* study, compound 8 also significantly suppresses tumor growth of breast cancer–initiating cells in mouse tumor model.

This method has the potential to be more efficient than conventional methods like virtual screening and high-throughput screening, but it is in the need for known inhibitors from which the fragment libraries are generated. For targets with many known inhibitors, such as kinases, *in silico* site-directed FBDD would likely be useful and produce many new leads. This drug design method can also be used for target without known inhibitors, if its homologies exist with sufficient known inhibitors. Fragment libraries built from these inhibitors still can be used to search for new inhibitors for the protein. An additional shortcoming is that ligand efficiency (LE) is unlikely to increase significantly during the fragment recombination, unless the binding modes of fragments can be greatly improved.
without increasing their sizes. One might not expect deconstruction of a potent lead to always derive high quality fragments if the synergy between two nonadjacent binding interactions is lost during the fragmentation. How reliably fragments can be recombined to maintain synergy will need to be further examined.

4.5 Conclusions

Recombination of two fragments from existing STAT3 dimerization inhibitors through in silico site-directed FBDD is an efficient and feasible method to design more potent STAT3 inhibitors. Linker selection was well studied so that the synthetic strategy was feasible; the linker maintained the initial binding modes of the fragments and enhanced the binding affinity of the fragments. The discovery of novel STAT3 inhibitors is a successful trial. The fluorescence polarization (FP) assay proved strong binding of 8 to STAT3 SH2 domain. The experimental ΔG value of 8 is -7.9 kcal/mol, which is comparable to the AutoDock4 computed ΔG value, -7.3 kcal/mol. Four out of five compounds have IC50 lower than 5µM for cancer cell line U2OS, among which, 8 has an IC50 ranging between 0.5-1.4µM in different cell lines. 8 and 11 have been demonstrated to significantly decrease STAT3 phosphorylation. Compound 8 shows a clear selectivity between P-STAT3 and P-STAT1 and also significantly suppresses tumor growth in vivo. The in silico site-directed FBDD strategy can be used to small molecule design to other drug targets.
4.6 Experimental section

4.6.1 FBDD.

General procedure of docking. Computational docking program AutoDock4 was used to dock all the existing inhibitors and our designed small molecules to predict their binding modes and approximate binding free energies to the STAT3 SH2 dimerization site. Compounds were docked using the Lamarckian Genetic Algorithm. The docking procedure involved the preparation of the ligand and macromolecule using the Schrödinger software. AutoDockTools was used to assign Gasteiger charges to the ligands. AutoGrid maps were then precomputed for all atom types in the ligand set. After 10 million energy evaluations were completed, all the resulting conformations of the ligands in the binding pocket of the macromolecule were clustered into groups according to their conformations with a root mean square deviation threshold of 1.5 Å. The most significant low energy clusters were identified and binding energies were evaluated.

4.6.2 Biological and biochemical assays

Cell Lines. Human breast cancer cell line (MCF-7), human sarcoma cell lines (RD2, RH30 and U2OS) and human medulloblastoma cell line (DAOY) were maintained in Dulbecco’s modified Eagle medium (DMEM) supplemented with 10% penicillin/streptomycin FBS and stored in a humidified 37°C incubator with 5% CO₂.
**STAT3 inhibitors.** 8, 11, and 13 were dissolved in sterile DMSO to make 20 mM stock solutions. Aliquots of the stock solutions were stored at -20°C.

**Cell Viability Assay.** Human sarcoma cell lines (RD2, RH30, U2OS) were seeded in 96-well plates at a density of 3,000 cells per well. The cells were incubated at 37°C for a period of 24 hours. Different concentrations of 8 (0.1-10 µM), 11 (0.1-20 µM), and 13 (0.1-10 µM) were added in triplicate to the plates in the presence of 10% FBS. 3-(4, 5-Dimethylthiazolyl)-2, 5-diphenyltetrazoliumbromide (MTT) was added to evaluate cell viability. The absorbance was read at 595 nm. IC\(_{50}\) values were determined using SigmaPlot 9.0 Software (Systat Software, Inc., San Jose, CA).

**Western Blot Analysis.** RH30 cells were treated with 8 (0.25-1 µM) or 11 (0.5-2.5 µM) or DMSO at 60% to 80% confluence in the presence of 10% FBS for 24 hours. The sarcoma cell line EW8 was treated with 8 (0.5-2.5 µM) or DMSO at 60% to 80% confluence in the presence of 10% FBS for 24 hours. MCF-7 cells and DAOY cells were treated with DMSO, 25-50 ng/ml of IL-6, 8 (0.25-1 µM) or 11 (0.5-2.5 µM), protein expressions of P-STAT3 (Tyr705) and STAT3 were tested. DAOY cells were treated with DMSO, 50 ng/ml of IFN-γ or 8 (1-5 µM), protein expressions of P-STAT3 (Tyr705) and STAT3 were tested.

Western blot procedure: The cells were harvested and lysed in cold radioimmunoprecipitation assay (RIPA) lysis buffer containing proteasome inhibitor
cocktail and phosphatase inhibitor cocktail, and were subjected to SDS-PAGE. Then they were transferred to PVDF membrane. Membranes were probed with primary antibodies (1:1,000) and horseradish peroxidase (HRP)-conjugated secondary antibody (1:10,000) (Cell Signaling Technology, Beverly, MA). Membranes were analyzed using Enhanced Chemiluminescence Plus reagents and scanned with the Storm Scanner (Amersham Pharmacia Biotech, Inc., Piscataway, NJ).

**IL-6 Induction of STAT3 Phosphorylation.** MCF-7 breast cancer cells were seeded in 10 cm plates and allowed to adhere overnight. The following day, the cells were serum-starved. The cells were then left untreated or were treated with 8 (0.25-1 μM) or 11 (0.5-2.5 μM) or DMSO. After 5 hours, the untreated and 8 or 11 treated cells were stimulated by IL-6 (25 ng/mL). The cells were harvested after 30 minutes and analyzed by Western blot. DAOY medulloblastoma cancer cells were seeded in 10 cm plates and allowed to adhere overnight. The following day, the cells were serum-starved. The cells were then left untreated or were treated with 8 (1-5 μM) or DMSO. After 2 hours, the untreated and 8 treated cells were stimulated by IL-6 (25 ng/mL). The cells were harvested after 30 minutes and analyzed by Western blot.

**IFN-γ Induction of STAT1 Phosphorylation.** DAOY medulloblastoma cancer cells were seeded in 10 cm plates and allowed to adhere overnight. The following day, the cells were serum-starved. The cells were then left untreated or were treated with 8 (1-5 μM) or DMSO.
µM) or DMSO. After 2 hours, the untreated and 8 treated cells were stimulated by IFN-γ (50 ng/mL). The cells were harvested after 30 minutes and analyzed by Western blot.

**Mouse Xenograft Tumor Model.** MDA-MB-231 human breast cancer cells (5 x 10⁶) were injected subcutaneously into the flank area of 6-week-old female athymic nude mice which were purchased from Harlan (Indianapolis, IN, USA). After tumor development, mice were divided into two treatment groups consisting of eight tumors per group: DMSO vehicle control, IP injection of 8 (5mg/kg). The inhibitors were formulated with Cremaphor, DMSO, and 5% dextrose to enhance delivery and limit toxicity encountered with DMSO alone as the mixing base. Tumor growth was determined by measuring length (L) and width (W) of the tumor every other day with a caliper. The tumor volume was calculated according to the formula: Tumor volume=0.5236×L×W². The treatment lasted for 21 days.

**Fluorescence Polarization (FP) Assay.** FP assay was used to analyze the ability of 8 to inhibit phosphopeptide binding to the STAT3 SH2 domain. STAT3 protein and peptides were provided by Dr. Pui-Kai Li’s lab. Screening was performed at approximately 37°C. STAT3 protein was used at 300 nM. The final concentration of buffer components used for all FP assays was 10 mM HEPES (pH 7.5), 1 mM EDTA, 0.1% Nonidet P-40, 50 mM NaCl, and 20% DMSO. Dithiothreitol was not added. 8 was diluted to a series of concentrations (1-200µM) from a 20 mM stock in DMSO. Proteins were incubated with 8 in Eppendorf tubes at room temperature for 60 min prior addition of the 5-
carboxyfluorescein labeled peptides. The final concentration of labeled peptide is 10 nM. The mixtures were transferred to a 384 Well Costar Black plate with each concentration repeated in three wells. The mixtures were equilibrated in the incubator at 37°C for at least 30 min. The FP readings were measured and the experiments were repeated three times. Inhibition curves with standard deviation were fitted and $K_i$ value was determined accordingly.

**Abbreviations Used:** HPLC, high performance liquid chromatography; DMSO, dimethyl sulfoxide; HEPES, (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid); EDTA, ethylenediaminetetraacetic acid; TLC, thin layer chromatography; IL-6, Interleukin 6; IFN-$\gamma$, Interferon $\gamma$. 
Chapter 5: Optimization of STAT3 anticancer inhibitors based on structure-activity study

5.1 Abstract

In our previous study, we have found several new compounds as potent STAT3 inhibitors by interrupting the dimerization process of phosphorylated STAT3. The cyclized compound 10 was found to be more potent than the ring-opening compound 9. In order to perform structure-activity relationship (SAR) study, more cyclized compounds and ring-opening compounds were synthesized to further verify the SAR results. The potency of LY5-1 was found to be improved or at least comparable to its ring-opening analog LY5 in two different cell lines. The IC$_{50}$ of LY5-1 was ranged from 0.45 to 1.46 μM for different cancer cell lines.

5.2 Introduction

Drug discovery is a sophisticated process, which incorporates several distinct phases each requiring a dedication of time and resources. Although more and more new drug design methods were developed, the conventional Structure-Activity Relationship (SAR) study still plays a very important role in drug design and development. In our previous study, we found the cyclized compound 10 (LY2) was more potent than the ring-opening compound 9 (LY7). More cyclized compounds and ring-opening compounds were
synthesized to conclude the SAR of these inhibitors (Figure 5.1). All of the compounds were tested by several different cancer cell lines. It was found that all the cyclized compounds were improved or at least comparable to the ring-opening compounds.

Figure 5.1 The chemical structures of the cyclized and ring-opening STAT3 inhibitors.
5.3 Results

5.3.1 SAR study

Utilizing *in silico* site-directed fragment-based drug design method, we previously found several novel potent inhibitors of STAT3 SH2 domain, which were active to multiple cancer cell lines. Two series of STAT3 inhibitors were synthesized, including cyclized compounds and ring-opening compounds. They are structurally different only in one portion.

5.3.2 Biological and Biochemical Assays

Cell Viability Assays. These compounds were tested in different cancer cell line. The IC$_{50}$ values for all of the compounds were listed in Table 5.1. These results showed that the cyclized compounds were more potent than the ring-opening compounds.

In this study, we synthesized two series of structurally related compounds to conclude the structure-activity relationship of these STAT3 inhibitors. From the comparison of the IC$_{50}$ in different cancer cells, the cyclized compounds were improved or at least comparable to the ring-opening compounds.
Table 5.1 IC$_{50}$ of the two series of STAT3 Inhibitors

<table>
<thead>
<tr>
<th>Compd</th>
<th>UW426</th>
<th>UW288-1</th>
<th>U2OS</th>
<th>cLogP</th>
</tr>
</thead>
<tbody>
<tr>
<td>LY2</td>
<td>5.0</td>
<td></td>
<td></td>
<td>1.19</td>
</tr>
<tr>
<td>LY7</td>
<td>&gt;5.0</td>
<td></td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td>LY5-1</td>
<td>0.45</td>
<td>1.46</td>
<td></td>
<td>-0.15</td>
</tr>
<tr>
<td>LY5</td>
<td>0.80</td>
<td>1.66</td>
<td></td>
<td>-1.04</td>
</tr>
<tr>
<td>LY3</td>
<td></td>
<td>1.4</td>
<td>3.29</td>
<td></td>
</tr>
<tr>
<td>LY8</td>
<td>&gt;5.0</td>
<td>&gt;5.0</td>
<td></td>
<td>1.72</td>
</tr>
<tr>
<td>LY6</td>
<td></td>
<td>2.5</td>
<td>2.18</td>
<td></td>
</tr>
<tr>
<td>LY9</td>
<td>&gt;5.0</td>
<td>&gt;5.0</td>
<td></td>
<td>1.29</td>
</tr>
<tr>
<td>LY17-R2</td>
<td>1.5</td>
<td>1.75</td>
<td></td>
<td>2.82</td>
</tr>
<tr>
<td>LY17-R3</td>
<td>2.3</td>
<td>&gt;5</td>
<td></td>
<td>1.93</td>
</tr>
</tbody>
</table>

5.4 Discussion

**SAR for site pTyr705** Albeit from the structure-based drug design aspect, the ring-opening compounds are more preferred than the cyclized compounds, since the free sulfonamide group can form two strong hydrogen bonds with the Arg609 in SH2 domain
of STAT3 protein, the final activity tests showed that the cyclized compounds are more potent. From the cLogP values, the cyclized compounds have better hydrophobicity than the ring-opening compounds. Thus, one explanation is that the cyclized compounds are more cell permeable than the ring-opening compounds, so that the real concentration of cyclized compounds inside cells are higher than that of the ring-opening compounds. The second possibility is that after cyclized compounds entering in the cytosol/nucleus, the cyclized compounds might be catalyzed by certain enzymes which move the equilibrium from the ring fused state to the ring opening state, and then the ring-opening compounds finally bind to the STAT3 SH2 domain.

**SAR for side pocket** The understanding of the features of side pocket is still far from enough till now, which impedes the drug development of STAT3 inhibitors. To better understand the side pocket, various fragments were tried to fit the binding site. From the *in vitro* results, the hydrophobic structure with polar substitutions might be a nice fit to the side pocket. For example, LY5/5-1 are more potent than LY2/7; LY3/8 are more potent than LY6/9. The highly planar structures of LY6/9 might be easier to form aggregation in cells. For LY17-R2 and LY17-R3, they are designed as prodrugs for STAT3 SH2 domain and they both have certain potency.
5.5 Conclusion

Two series of STAT3 inhibitors were synthesized to investigate the structure-activity relationship of the ring-fused and ring-opening compounds. The IC$_{50}$ values of these compounds showed that the cyclized compounds improved or at least comparable to the ring-opening compounds, which provides us useful information for further drug discovery of STAT3 inhibitors.

5.6 Experimental section

5.6.1 Chemistry.

Refer to Chapter 5.

5.6.2 Biological Assays.

Cell Lines. Human medulloblastoma cell lines (UW426, UW288-1) and human sarcoma cell line (U2OS) were maintained in Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% penicillin/streptomycin FBS and stored in a humidified 37 °C incubator with 5% CO2.

STAT3 Inhibitors. Compounds were dissolved in sterile DMSO to make 20 mM stock solutions. Aliquots of the stock solutions were stored at −20 °C.
Cell Viability Assay. Human medulloblastoma cell lines (UW426 and UW288-1) were seeded in 96-well plates at a density of 3000 cells per well. The cells were incubated at 37 °C for a period of 72 h. Human sarcoma cell line (U2OS) was seeded in 96-well plates at a density of 3000 cells per well. The cells were incubated at 37 °C for a period of 24 h. Different concentrations of LY compounds (0.1-20 μM) were added in triplicate to the plates in the presence of 10% FBS. 3-(4,5-Dimethylthiazolyl)-2, 5-diphenyltetrazoliumbromide (MTT) was added to evaluate cell viability. The absorbance was read at 595 nm. IC50 values were determined using SigmaPlot 9.0 Software (Systat Software, Inc., San Jose, CA).
Chapter 6: Computational design of STAT3 anticancer inhibitors using multiple ligand simultaneous docking method

6.1 Abstract

A improved *in silico* fragment-based drug design method, Multiple Ligand Simultaneous Docking (MLSD)\textsuperscript{17}, was used to discover a new series of STAT3 inhibitors by interrupting the dimerization process of phosphorylated STAT3. Fragments library was collected from the FDA approved drugs in market and in clinical trials and from known STAT3 inhibitors. The best fragment combination resulting from the MLSD calculation were linked together and then synthesized and tested. LY17 was found to inhibit STAT3 protein with $K_D$ value $2.42 \pm 0.26 \, \mu M$ by microscale thermophoresis test. The IC\textsubscript{50} of LY17 was ranged from 1.15 to 5.11\,\mu M for different cancer cell lines.

6.2 Introduction

Drug discovery still suffers with low efficiency and high risk. A successful drug needs to meet many requirements, such as high potency, good bioavailability, suitable pharmacokinetic and pharmacodynamic properties, high safety, few side effects, and low cost for large scale production, etc. Dissatisfaction in any of the above mentioned attributes will directly cause the failure of drug discovery, so that drug discovery is
always described as searching a needle in haystacks. Nowadays, due to the highly advanced computational techniques, large scale calculation can be realized. *In silico* fragment-based drug design as one of the most popular computational drug design methods arouses more and more attention to improve the efficiency of drug discovery. Multiple Ligand Simultaneous Docking (MLSD) is a modified *in silico* FBDD method which simulates the interplay of multiple molecules binding to the protein binding site(s) in order to compensate the nonspecific binding of conventional single fragment. In order to search for druggable STAT3 inhibitors, herein, we utilized MLSD method and searched for suitable combinations of fragments which can fit the binding sites of STAT3, site pTyr705 and side pocket. From previous studies, the regularly found drug scaffolds, ring systems and the small fragments, were identified and ranked according to the frequencies of their occurrence in drug molecules\(^{54,55}\). We further categorized these scaffolds and fragments into six fragment libraries according to their physical properties, polar_uncharged_nonaromatic fragments, polar_uncharged_aromatic fragments, polar_charged_amines, polar_charged_acids, nonpolar fragments, and large and bulk fragments, among which, polar fragment libraries were used in this study. Two fragments were docked to the STAT3 SH2 domain simultaneously and the best posed and scored fragment couples were linked together by distance measurement. A new series of STAT3 inhibitors were discovered, represented by LY17.

6.3 Results
The constitutive activation of STAT3 signaling pathway is triggered by the dimerization of phosphorylated STAT3 protein, followed by nuclear translocation, DNA binding and cell replication. The inhibition of STAT3 dimerization has been proved to be a viable approach to block the activation of STAT3 pathway. The binding cleft of STAT3 protein contains two subpockets, site pTyr705 and side pocket. Fragments binding to these two subpockets can be linked to inhibit STAT3 dimerization. In our previous study\textsuperscript{53,56}, naphthalene-5,8-dione-1-sulfonamide (1), derived from STAT3 inhibitor LLL12 (Figure 6.1), can bind to site pTyr705 most tightly, so that it is chosen as the fragment fit to site pTyr705. The other fragment binding to the side pocket was selected from reorganized fragment libraries (963 fragments in total) by properties, 280 polar\_uncharged\_nonaromatic fragments, 532 polar\_uncharged\_aromatic fragments and 61 polar\_charged\_amines, 90 polar\_charged\_acids, built on FDA approved drugs in market and in clinical trials.

![Figure 6.1 Chemical structures of fragment 1 and LLL12.](image-url)
6.3.1 Multifragment Simultaneous Docking Screening

The crystal structure of STAT3 SH2 domain (PDB: 1BG1) was used as the receptor structure for docking. Fragment 1 and each single fragment of the 963 fragments were grouped together and docked to the receptor employing the MLSD docking program. Two fragments as one group were used to probe binding subpockets of STAT3 SH2 domain. Systematic multifragment virtual screening was ranked by the predicted binding energies. The docked fragment groups with a predicted binding energy less than -8.5 kcal/mol were considered for further visual inspection of binding modes (Appendix B).

6.3.2 Fragment Linkage

As fragment 1 favors binding to site pTyr705, linkers are selected according to the available synthetic approaches to connect fragment 1 to the other fragment. The linking point was determined based on the distance measurement. The nearest carbon on the other fragment ring to C6 on fragment 1 was added with a primary amine group, so that two fragments can be linked by the C-N coupling reaction (Figure 6.2).
6.3.3 Hit Selection

The linked compounds were docked to the STAT3 SH2 crystal structure and ranked by their docking energies. Compounds which can reposition to the initial docking modes of two separate fragments were selected, and their docking energies were ranged from -8.5 ~ -9.5 kcal/mol. The structures of selected compounds (Figure 6.3) and their docking energies are listed in Table 6.1. The reposition of selected compounds is listed in Appendix C.

By the comparison of the docked conformations before/after linkage, the selected compounds can maintain the interactions observed in multifragments and even improved their initial binding modes. Limited by the availability of fragments as starting materials for organic synthesis, the selected compounds for further synthesis and biological tests were LY10, LY13, LY16 and LY17.
Table 6.1 Predicted compounds and their selectivity to STAT3 and STAT1.

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Binding energy to STAT3 (kcal/mol)</th>
<th>Binding energy to STAT1 (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLL12</td>
<td>-7.1</td>
<td>NB(^a)</td>
</tr>
<tr>
<td>LY-5</td>
<td>-7.3</td>
<td>NB</td>
</tr>
<tr>
<td>LY-9</td>
<td>-9.5</td>
<td>-7.8</td>
</tr>
<tr>
<td>LY-10</td>
<td>-9.3</td>
<td>NB</td>
</tr>
<tr>
<td>LY-11</td>
<td>-9.2</td>
<td>NB</td>
</tr>
<tr>
<td>LY-12</td>
<td>-9.1</td>
<td>NB</td>
</tr>
<tr>
<td>LY-13</td>
<td>-9.0</td>
<td>NB</td>
</tr>
<tr>
<td>LY-14</td>
<td>-9.0</td>
<td>NB</td>
</tr>
<tr>
<td>LY-15</td>
<td>-8.5</td>
<td>NB</td>
</tr>
<tr>
<td>LY-16</td>
<td>-8.5</td>
<td>NB</td>
</tr>
<tr>
<td>LY-17</td>
<td>-8.5</td>
<td>NB</td>
</tr>
<tr>
<td>LY-18</td>
<td>-8.5</td>
<td>NB</td>
</tr>
</tbody>
</table>

\(^a\)NB means there is no binding between compounds and STAT1.
6.3.4 Biological and Biochemical Assays

**Cell Viability Assays.** Initially, IC\textsubscript{50} values for all of the compounds were tested in UW426 cell line. Then selected potent compounds were further tested in two other UW288-1 and BKPC3 cell lines (Table 6.2). LY17 and LY13 were found to be very potent compounds to kill different cancer cells.
Table 6.2 IC$_{50}$ of designed STAT3 Inhibitors.

<table>
<thead>
<tr>
<th>Compd</th>
<th>IC$_{50}$ values (µM) vs. Cell lines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UW426</td>
</tr>
<tr>
<td>LY13</td>
<td>1.49</td>
</tr>
<tr>
<td>LY10</td>
<td>&gt;5.0</td>
</tr>
<tr>
<td>LY17</td>
<td>1.16</td>
</tr>
</tbody>
</table>

LY17 Inhibits STAT3 Phosphorylation and Induces Apoptosis in Human Medulloblastoma Cancer Cells. To examine the inhibition of STAT3 phosphorylation, Western blot assays were performed to detect the abundance of phosphorylated STAT3 (P-STAT3) after UW426 medulloblastoma cell line was treated with LY17 (1-2.5 µM) or LY17-R2 (1-2.5 µM) for 12 h. STAT3 phosphorylation was reduced in a dose-dependent manner, with LY17 almost completely inhibiting Tyr705 phosphorylation at 2.5 µM. The expression of total STAT3 was not changed, indicating that the decrease of P-STAT3 was not due to a constitutional decrease of total STAT3 expression. LY17 was not found to inhibit phosphorylation of the kinase ERK1/2 and Akt. The inhibition of STAT3 phosphorylation by LY17 seems to be consistent with the induction of apoptosis as evidenced by the presence of cleaved caspase 3 (Figure 6.4).
Figure 6.4 Human Medulloblastoma cell line UW426 was treated with LY17 (1-2.5 µM) and LY17-R2 (1-2.5 µM) for 12 h, then whole-cell extracts were prepared and phospho-STAT3 was detected by Western blot assay, revealing a decrease in STAT3 phosphorylation.

**LY17 Inhibits STAT3 Phosphorylation Induced by IL-6.** IL-6 can induce the activation of STAT3. The Daoy medulloblastoma cancer cell line was used to determine the ability of LY17 to inhibit IL-6 induced STAT3 phosphorylation. In control experiments, IL-6 elevated STAT3 phosphorylation in Daoy cells, while LY17 and
LY17-R2 blocked the stimulation and abrogated phosphorylated STAT3 at low concentrations (Figure 6.5).

Figure 6.5 LY17 and LY17-R2 inhibit IL-6 induced STAT3 phosphorylation in Daoy cancer cells. The Daoy cells were serum-starved overnight, then left untreated or treated with LY17 or LY17-R2 (1-2.5 µM) for 2h, followed by stimulation by IL-6 (50 ng/mL). The cells were harvested at 30 minutes and analyzed by Western blot assays.

The Dissociation Constant of LY17 to STAT3 SH2 Domain Measured by Microscale Thermophoresis (MST). In the MST-Experiment we have kept the concentration of the NanoTemper-labeled STAT3 constant (30 nM), while the concentration of the non-labeled binding partner (LY17) was varied between 1.52 nM-50,000 nM. After a short incubation the samples were loaded into MST standard glass capillaries and the MST-
analysis was performed using the Monolith.NT115. All concentrations are in nM. A $K_D$ of $2,420 \text{ nM} \pm 256 \text{ nM}$ was determined for this interaction (Figure 6.6).

![Figure 6.6 Microscale thermophoresis (MST) tests on LY17 as a modified STAT3 inhibitor.](image)

6.4 Discussion

In this study, we utilized the newly developed fragment-based drug design approach, Multiple Ligand Simultaneous Docking method, using two fragments together as a group to probe the chemical space of STAT3 SH2 domain. The docked results after the
competition between two fragments to several binding sites showed us a blueprint of
fragment binding and especially provided us the rationales to assign suitable fragment to
specific binding site. This method is used to differentiate the nonspecific binding of
conventional single fragment to adjacent binding sites.

Secondly, regioisomers in many cases possess different activity, toxicity or druggability.
MLSD is a good option to help determine the stereochemistry of designed compounds. In
MLSD, in order to link the two fragments, the stereochemistry linking at either ortho,
meta or para position can be predicted by the distance measurement based on the docked
3D structures of the fragment combination.

Since building blocks of the existing drugs and clinical drugs share many interesting
properties, it was reported that about 50 fragments can cover both 52.6% drug scaffolds
of 1240 FDA-approved drugs and 48.6% fragments of FDA-approved drugs and
experimental drugs in different phases of clinical trials from several drug databases (6932
entries)\textsuperscript{54}. In addition, fragments from existing drugs are considered to be efficient and
safe compared to random fragments. So fragment libraries built on existing drugs are
used in our study aiming to design more druggable STAT3 inhibitors.

From the structural analysis, the side pocket binding site appeals to be primarily
hydrophobic with several polar residues, so that hydrophobic fragments with polar
substitution will be a good fit to the site pocket. In this study, the polar fragment libraries were used together with fragment 1 for scaffold hopping.

Fragments were linked together based on distance measurements, which were further repositioned to SH2 domain. Limited by the feasibility of organic synthesis, C-N coupling reaction was chosen to link the two selected fragments together. So the key step of the MLSD study is the linker selection, which determines the synthetic feasibility of designed compounds.

6.5 Conclusion

Seven new STAT3 inhibitors were discovered by a modified fragment-based drug design method, MLSD. Among these inhibitors, LY17 was proved to bind to STAT3 protein with $K_D$ value $2,420 \text{ nM} \pm 256 \text{ nM}$ by MST tests. The $IC_{50}$ of LY17 was ranged from 1.15 to $5.11 \mu\text{M}$ for different cancer cell lines. In addition, LY17 has been demonstrated to significantly decrease STAT3 phosphorylation/activation.

6.6 Experimental section

6.6.1 MLSD

The fragment libraries were reorganized by properties using Canvas in Schrödinger$^{57}$. Computational docking program MLSD$^{17}$ was used to dock all the fragment groups (fragment 1 and fragment from the polar fragment libraries) to predict their binding
modes and approximate binding free energies to the STAT3 SH2 dimerization site. The particle swarm optimization (PSO) method was used to screen out the good fragment combinations ranked by docking energies. Linked compounds were redocked to STAT3 SH2 screened by Lamarckian genetic algorithm (LGA) and ranked by binding energies and binding modes to generate hits for synthesis and cell assays. The docking procedure involved the preparation of the ligand and macromolecule using the Schrodinger software\textsuperscript{58,59}. AutoDockTools was used to assign Gasteiger charges to the ligands. AutoGrid maps were then precomputed for all atom types in the ligand set. After 10 million energy evaluations were completed, all the resulting conformations of the ligands in the binding pocket of the macromolecule were clustered into groups according to their conformations with a root-mean-square deviation threshold of 1.5 Å. The most significant low energy clusters were identified and binding energies were evaluated.

\textit{6.6.2 Biological Assays and biochemical assays.}

Cell Lines. Human medulloblastoma cell lines (UW426, UW288-1 and Daoy) and human pancreatic cancer cell line (BkPC-3) were maintained in Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% penicillin/streptomycin FBS and stored in a humidified 37 °C incubator with 5% CO\textsubscript{2}.

STAT3 Inhibitors. Newly designed compounds were dissolved in sterile DMSO to make 20 mM stock solutions. Aliquots of the stock solutions were stored at \(-20 \, ^\circ\text{C}\).
Cell Viability Assay. Human medulloblastoma cell lines (UW426, UW288-1 and BkPC-3) were seeded in 96-well plates at a density of 3000 cells per well. The cells were incubated at 37 °C for a period of 24 h. Different concentrations of LY17 (0.1-10 μM), LY17-R2 (0.1-20 μM), and LY13 (0.1-10 μM) were added in triplicate to the plates in the presence of 10% FBS. 3-(4,5-Dimethylthiazolyl)-2,5-diphenyltetrazoliumbromide (MTT) was added to evaluate cell viability. The absorbance was read at 595 nm. IC$_{50}$ values were determined using SigmaPlot 9.0 Software (Systat Software, Inc., San Jose, CA).

Western Blot Analysis. UW426 cells were treated with LY17 (1-2.5 μM) or LY17-R2 (1-2.5 μM) or DMSO at 60−80% confluence in the presence of 10% FBS for 24 h. Daoy cell line was treated with DMSO, or 50 ng/mL of IL-6, or IL-6/LY17 (1-2.5 μM) or IL-6/LY17-R2 (1-2.5 μM), protein expressions of P-STAT3 (Tyr705) and STAT3 were tested.

Western blot procedure: The cells were harvested and lysed in cold radioimmunoprecipitation assay (RIPA) lysis buffer containing proteasome inhibitor cocktail and phosphatase inhibitor cocktail, and were subjected to SDS-PAGE. Then they were transferred to PVDF membrane. Membranes were probed with primary antibodies (1:1,000) and horseradish peroxidase (HRP)-conjugated secondary antibody (1:10000) (Cell Signaling Technology, Beverly, MA). Membranes were analyzed using Enhanced Chemiluminescence Plus reagents and scanned with the Storm scanner (Amersham Pharmacia Biotech, Inc., Piscataway, NJ).
IL-6 Induction of STAT3 Phosphorylation. Daoy cancer cells were seeded in 10 cm plates and allowed to adhere overnight. The following day, the cells were serum-starved. The cells were then left untreated or were treated with LY17 (1-2.5μM) or LY17-R2 (1-2.5 μM) or DMSO. After 2 h, the untreated and treated cells were stimulated by IL-6 (50 ng/mL). The cells were harvested after 30 min and analyzed by Western blot.

Microscale Thermophoresis (MST). A titration series of up to 16 dilutions of the inhibitor LY17 was prepared. Here the concentration of the fluorescently labeled STAT3 protein was kept constant at 10-20 nM and the concentration of the titrant was varied. Therefore 2-20 μM of proteins of interest were labeled with a fluorescent dye (NT-647, NT-532 or NT-488) using Monolith NT™ Protein Labeling Kits (amine or cysteine reactive). The labeling procedure and the subsequent removal of free dye were performed within 45 min. In the dilution series the highest concentration was chosen to be 20-fold higher than the expected K_D. 10 μl of the serial dilution of the non-labeled molecule were mixed with 10 μl of the diluted fluorescently labeled molecule. Mixed samples were loaded into glass capillaries and the MST-analysis was performed using the Monolith.NT115.
Chapter 7: Regioselective one-pot C-N coupling of substituted naphthoquinones: selective intramolecular ring fusion of sulfonamide

(A manuscript submitted to the Tetrahedron)

7.1 Abstract

The first one-pot copper-catalyzed highly regioselective C-N bond formation between aryl/alkyl amines and sulphonamide substituted naphthoquinones was accomplished. Facile chemoselective routes obtained diverse ring-opening 6-amino/anilino-naphthalenedione-1-sulfonamides and ring-fused 6-amino/aniline-5H-naphth[1,8-cd]isothiazol-5-one,1,1-dioxides with great functional group tolerance. Stereochemistry was confirmed by NMRs.

7.2 Introduction

Naphthoquinone is a prevalent scaffold found in a wide variety of natural products and drug candidates, for example, vitamin K₁ and K₂, JNK inhibitors plumbagin, shikonin and isothiazoloanthrone (1)\textsuperscript{60,61}, antitumor agent aminonaphthoquinones (2)\textsuperscript{62-64}, and antifungal candidate aminonaphthalene-1,4-diones (3)\textsuperscript{65}, etc (Figure 7.1). The sulfonamide chemical moiety is also present in various therapeutic drugs, which contains both hydrogen bond donors (NH₂) and acceptors (O=S=O). One sulfonamide group can
form as many as seven hydrogen bonds and it tends to form strong interactions with certain protein targets, thus it becomes a commonly used functional group in drug design and discovery\textsuperscript{66}. In our study, the hybrid of napthoquinone and sulfonamide group especially provides a novel series of compounds as very potent anticancer agents\textsuperscript{67}.

![Figure 7.1 Structures of naphthalene-1,4-dione derivatives.](image)

Lots of efforts were spent on the synthesis of amines and anilines with unsubstituted 1,4-naphthoquinone to give 2-amino/anilino-naphthalene-1,4-dione\textsuperscript{68-70}. Once the 1,4-naphthoquinone is substituted with groups (-OH, -NH\textsubscript{3}, -Cl, -SO\textsubscript{2}NH\textsubscript{2}, -OMe, etc.), C-N coupling happening at position 2 or 3 will lead to two different regio-isomers, which raises the issue of regioselective synthesis. Two general strategies for the preparation of regioselective aminonaphthoquinone derivatives have been reported: first, nucleophilic substitution reactions, for example, mono-halo (Br) or di-halo (2Cl/2Br) substituted
naphthoquinones react with amines to produce regioisomers\textsuperscript{71-74}; however, the methods usually suffer from low yields and tedious purifications. Extra steps are required to prepare the halo substituted naphthoquinones, which limited the choices of substituted groups of naphthoquinones. So the approach is not applicable to sensitive groups, such as sulfonamide, which cannot tolerate to bromination or chlorination. Second, oxidative coupling reactions, for example, gold (III) salt is used to catalyze the 1,4-nucleophilic addition of amines to unsubstituted 5,8-quinolinedione\textsuperscript{75}. To synthesize regioselective substituted naphthoquinones, more efficient and less costly approach is still necessary to develop. To the best of our knowledge, the coupling reaction between sulfonamide-substituted naphthoquinone derivatives and amines has never been reported so far. Although the condensation between sulfonamide and carbonyl group can be separately catalyzed by concentrated H\textsubscript{2}SO\textsubscript{4}\textsuperscript{76}, we herein report one-pot reactions with great functional group tolerance, incorporating highly regioselective coupling between naphthoquinosulfonamide and aryl/alkyl amines, and chemoselective synthesis of ring-fused or ring-opening naphthoquinone derivatives.
Table 7.1 Optimization of reaction conditions

![Chemical structure](image)

**Reaction conditions:** naphthalene-5,8-dione-1-sulfonamide 4 (237 mg, 1 mmol), aniline (0.36 mL, 1.2 mmol), Cu(OAc)$_2$.H$_2$O (40 mg, 0.2 mmol).

<table>
<thead>
<tr>
<th>Entry</th>
<th>Catalyst</th>
<th>Solvent</th>
<th>S1/S2$^a$ (v/v,%)</th>
<th>T(°C)</th>
<th>Yield$^b$ (%) (2a/3a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cu(OAc)$_2$.H$_2$O</td>
<td>AcOH</td>
<td>100</td>
<td>60</td>
<td>26/20</td>
</tr>
<tr>
<td>2</td>
<td>Cu(OAc)$_2$.H$_2$O</td>
<td>AcOH</td>
<td>100</td>
<td>80</td>
<td>71/trace</td>
</tr>
<tr>
<td>3</td>
<td>Cu(OAc)$_2$.H$_2$O</td>
<td>AcOH/H$_2$O</td>
<td>90/10</td>
<td>80</td>
<td>27/30</td>
</tr>
<tr>
<td>4</td>
<td>Cu(OAc)$_2$.H$_2$O</td>
<td>AcOH/H$_2$O</td>
<td>50/50</td>
<td>80</td>
<td>11/15</td>
</tr>
<tr>
<td>5</td>
<td>Cu(OAc)$_2$.H$_2$O</td>
<td>THF</td>
<td>100</td>
<td>65</td>
<td>37/10</td>
</tr>
<tr>
<td>6</td>
<td>Cu(OAc)$_2$.H$_2$O</td>
<td>EtOH</td>
<td>100</td>
<td>77</td>
<td>41/8</td>
</tr>
<tr>
<td>7</td>
<td>CeCl$_3$</td>
<td>AcOH</td>
<td>100</td>
<td>rt</td>
<td>35/13</td>
</tr>
<tr>
<td>8</td>
<td>CeCl$_3$</td>
<td>THF</td>
<td>100</td>
<td>rt</td>
<td>21/16</td>
</tr>
<tr>
<td>9</td>
<td>CeCl$_3$</td>
<td>EtOH</td>
<td>100</td>
<td>rt</td>
<td>32/11</td>
</tr>
</tbody>
</table>

$^a$The volume percent of different solvent systems. $^b$The crude products obtained were purified through a flash column of silica gel, then recrystallized from EtOH/H$_2$O or THF/H$_2$O. The yields given in the table are for the first crop of crystals.
Figure 7.2 Scope of ring-fused 6-amino/aniline-5H-naphth[1,8-cd]isothiazol-5-one,1,1-dioxides$^{ab}$. $^a$ Reaction conditions: naphthalene-5,8-dione-1-sulfonamide (237 mg, 1 mmol), alkyl/aryl amine (1.2 mmol), Cu(OAc)$_2$.H$_2$O (40 mg, 0.2 mmol) in glacial acetic acid (5 mL) at 80-100 °C, air. Reaction was monitored by TLC detection (ESI†). $^b$Isolated yields.

Recent studies on the oxidative coupling reactions of unsubstituted aminonaphthoquinone were reported that the use of Ce(III) salts in ethanol$^{62}$ and Cu(II) salts in AcOH$^{77}$ both in the presence of oxygen resulted in good yields. These methodologies can avoid the prehalogenation of the napthoquinones and provide possible alternative routes for the coupling reactions between napthoquinosulfonamide and amines. To study concise
routes on the regioselective coupling of naphthalene-5,8-dione-1-sulfonamide with aniline (Table 7.1), preliminary attempts using either Cu(OAc)$_2$.H$_2$O as catalyst in AcOH at 60°C (entry 1) or CeCl$_3$ as catalyst in ethanol (entry 9) at room temperature were performed, resulting the C-N direct amination of the desired C6-aminated products 3a and another ring-fused product 2a, without C7-aminated regio-isomers. Initially, a screening of reaction temperature was done, and the yield of 2a increased to 71% at temperature 80°C (entry 2). Then an effort to increase the yield of 3a was made. In order to avoid the dehydration of the sulfonamide group, different ratios of water were added to the reaction solution. From the results (entries 3-4), water was not a favorable solvent for the synthesis of 2a, however, it did improve the yield of 3a. The yield of 3a was increased to 30% in 10% water/AcOH co-solvent, 80 °C. In order to investigate the best solvent for the coupling reaction, other solvents were tried, including THF and EtOH (entries 5,6 and 8,9), turned out that glacial acetic acid was the best choice. As Ce(III) salt was also reported as an efficient catalyst for oxidative coupling, we tried to compare the effects of Ce(III) salts and Cu(II) salts. Both catalysts have their own advantages. Reactions catalyzed by Cu(OAc)$_2$.H$_2$O gave a higher yield and less byproducts than catalyzed by CeCl$_3$. In the meanwhile, catalyst CeCl$_3$ facilitated the reaction at a relative lower temperature, such as room temperature, which provides an opportunity for reactions with thermo-sensitive reactants/reagents.

Since none of the above conditions led to a majority of the desired compound 3a and reaction temperature did not appear to be the key reason to cause the cyclization of the
sulfonamide group with the adjacent ketone group, our initial attempts focused on hydrolysis of the C-N double bond to free the sulfonamide group from the fused ring. After a screening of both basic and acidic conditions, the fused ring was found to be opened in a refluxed TFA/EtOH co-solvent, yield 67% (ESI†). Then inspired by the hydrolysis reaction of 2a, a modified synthetic method of 3a was built. Compound 1, aniline and catalyst Cu(OAc)₂.H₂O reacted in a co-solvent system (TFA/AcOH v/v:1/10) at 60°C for 3h, resulting in pure product 3a, yield 84%, without the production of the ring-fused compound 2a. Therefore, two distinct reaction conditions were found to chemoselectively lead to either the ring-fused or the ring-opening products.

With the optimized conditions established, firstly the reactions of various alkyl/aryl amines to synthesize the ring-fused product were explored (Figure 7.2). A range of amines were tolerated under these conditions. For hetero-arylamines, such as pyridin-3-amine, the reaction gave the fused-ring product (2b) in moderate yield in pure glacial acetic acid. For aniline substituted with electron-withdrawing groups, such as 2-chloro-4-nitroaniline, higher energy was required to cross the transition state to synthesize 2c. The reaction did not happen until the temperature was increased to 100 °C. Intriguingly, both zwitterions, carboxyl group substituted alkyl and aryl amines, can highly selectively react with naphthalene-5,8-dione-1-sulfonamide to synthesize 2e, 2f. A higher temperature and longer time was required. It is also noteworthy that protected amine by an ester bond, for instance, protected by benzyl formate (Cbz) group, can also smoothly react with naphthalene-5,8-dione-1-sulfonamide to give corresponding product 2g in pure
glacial acetic acid solvent condition. So the tolerance of Cbz protected amines makes it possible to selectively react with substrates with multiple amines. In all, the reaction can tolerate different functional groups, including electron-donating/withdrawing groups, alkyl/aryl amines, carboxyl groups, and ester bonds, which provides an opportunity to synthesize various coupled products.

Figure 7.3 Scope of ring-opening 6-amino/aniline-naphthalene-5,8-dione-1-sulfonamides\(^ac\). \(^a\)Reaction conditions: naphthalene-5,8-dione-1-sulfonamide (237 mg, 1 mmol), alkyl/aryl amine (1.2 mmol), Cu(OAc)\(_2\)·H\(_2\)O (30 mg, 0.15 mmol) in co-solvent TFA/glacial acetic acid (0.05-0.5 mL/5 mL v/v) at certain temperature, air. Reaction was monitored by TLC detection (ESI\(^+)\). \(^b\)3g was synthesized in co-solvent water/glacial acetic acid (0.5 mL/5mL v/v). \(^c\)Isolated yields.
Next, the effective conditions were extended to diverse ring-opening products as illustrated in Figure 7.3 (3a-g). For the TFA/AcOH co-solvent condition, alkyl and aryl amines with various substituents can be converted to the desired products in good to excellent yields. Subsequently, a screening of the ratios of TFA/AcOH showed that the higher the percentage of TFA, the lower the yield. To minimize the ring-fused product, but maximize the yield, a volume ratio of TFA/AcOH, 1/10, is suitable for most of the products. For hetero-arylamines, such as pyridin-3-amine, the ratio can be decreased to 1% TFA in AcOH to increase the yield. For aniline substituted with electron-withdrawing groups, such as 2-chloro-4-nitroaniline, reaction temperature was increased to 80 °C. For carboxyl group substituted alkyl and aryl amines, the co-solvent was heated to reflux to completely consume the reagent naphthalene-5,8-dione-1-sulfonamide. However, the reaction of Cbz protected amine with naphthalene-5,8-dione-1-sulfonamide did not give corresponding product 3g in the co-solvent TFA/AcOH condition and several by-products were observed. Finally, 3g was synthesized in a co-solvent system of 10% water in glacial acetic acid at 80 °C.
Based upon the above results, a plausible mechanism is proposed in Scheme 7.1. One possible mechanism, the oxidative coupling reaction initially involves a Michael addition of alkyl/aryl amines to naphthalene-5,8-dione-1-sulfonamide, catalyzed by Cu(II) salt. Subsequently, the resulting copper hydroquinone complex gives free radical with the aid of the acetic acid and rearranges electrons and double bonds, and then the desired ring-opening product was formed. The resulting Cu(I) salt can be rescued by oxygen in air, and then re-enter the oxidation coupling. Control experiments were conducted. When the reaction was protected by nitrogen, trace amount (yield <5%) of desired products were synthesized, which indicated that the molecular oxygen acts as the oxidant. When the reaction was set up without catalyst Cu(II) salt, no desired products were found. So Cu(II) can greatly facilitate the reaction and assist to cross the transition state energy
threshold. When the reaction was set up without alkyl/aryl amines, no cyclized products were found. So the condensation is triggered by the oxidative coupling (ESI†). It is also possible that the cyclization helps to determine the regioselectivity of the reaction. And a stronger acidic medium TFA moves the equilibrium from the ring fused state to the ring opening state.

To identify the stereochemistry of the regioisomer, several one-dimensional and two-dimensional NMRs were performed, including $^1$HNMR, $^{13}$CNMR, DEPT-90, HMBC, HSQC, $^1$H$^1$HCOSY (ESI†). The amination coupling was identified at position 6 but not 7.

7.3 Conclusions

We have reported a novel regioselective Cu(II)-catalyzed direct amination of naphthalene-5,8-dione-1-sulfonamide through C-N bond formation with excellent functional group tolerance, which successfully avoids the halogenation of the naphthoquinone ring and protecting the sulfonamide group. This study also provides a selective method to synthesize both ring-fused 6-amino/aniline-$5H$-naphth[1,8-cd]isothiazol-5-one,1,1-dioxides and ring-opening 6-amino/anilino-naphthalene-5,8-dione-1-sulfonamide products. This protocol can not only serve as a new avenue for amination coupling of sensitive functional groups substituted naphthoquinone, but also
afford a facile approach for the synthesis of various bioactive molecules. The possible mechanistic pathway was also proposed on the basis of control experiments.

7.4 General Information

Reactions were monitored by thin layer chromatography with fluorescent silica coated aluminum sheets. Melting points (mp) were determined on a Thomas Hoover capillary melting point apparatus. $^1$H NMR and $^{13}$C NMR spectra were recorded on Bruker Avance 300 MHz and 400 MHz NMR spectrometers (Billerica, MA) at room temperature using DMSO as solvent. Chemical shifts (δ) are reported in ppm with the solvent resonance as an internal standard (DMSO-$d_6$: δ = 2.49 ppm). Abbreviations for signal couplings are: s, singlet; d, doublet; t, triplet; m, multiplet. Chemical shifts of $^{13}$C NMR were reported in ppm with the solvent as the internal standard (DMSO-$d_6$: δ = 40 ppm, (CD$_3$)$_2$CO-$d_6$: δ = 29.84 ppm). Electron spray ionization (ESI) was used as the ion source for high resolution mass measurement. Analytical HPLC was carried out with a Gemini 5μ C18 110A Column (250 x 4.6 mm) supplied by Phenomenex Inc. CA, USA, eluting with acetonitrile and water. The peaks were detected by UV light of irradiation (254 nm).

7.5 Materials
All reactions were carried out under air atmosphere. All reagents and solvents were obtained from commercial suppliers and used without further purification. The solvents and reagents used in the present study were purchased from commercial suppliers and were used as received. Silica gel was purchased from Sigma-Aldrich Chemical Co. (Milwaukee, WI). Catalysts Cu(OAc)$_2$.H$_2$O and CeCl$_3$, alkyl/aryl amines are commercially available. Naphthalene-5,8-dione-1-sulfonamide 1 was prepared according to previous literature$^{56}$, as shown in Scheme 7.1. Benzyl 4-(2-aminophenyl) piperazine-1-carboxylate was prepared as shown in Scheme 7.2. Unless otherwise indicated, all other reagents and solvents were obtained from commercial suppliers and used as received.

![Scheme 7.2 Preparation of compound 1. Reaction Conditions: (a) NH$_3$.H$_2$O, Acetone, 0 °C to rt; (b) CrO$_3$, AcOH, 60 °C.](image)

Synthetic procedure: Naphthalenesulfonylchloride (16.8 g, 74.3 mmol) was dissolved in acetone (100 mL) and was stirred at 0 °C for 5 minutes. Ammonium hydroxide (100 mL) was dropped into the above mixture and stirred at room temperature for 3 hours.
Precipitated white crystals were filtered then acetone was removed at reduced pressure. The residue was washed in ethylacetate (3×10 mL), producing a white solid powder which was used without further purification, yielding naphthalenesulfonamide (15.9 g, 90.2%); mp (147-149 °C). Naphthalenesulfonamide (500 mg, 2.41 mmol) was dissolved in slowly warming glacial acetic acid (5.0 mL). The mixture was heated to 90 °C and chromium trioxide (1.08 g, 10.85 mmol), which was dissolved in a mixture of water and glacial acetic acid (3 mL, 1:1 v/v), was added to the mixture solution. The above solution was stirred under reflux for 18 minutes and quenched with iced water (50 mL). The solution was cooled to 0 °C and stirred 1h at room temperature. The precipitated yellow powder was filtered and the remaining solution was extracted with ether (3×100 mL). The organic layer was collected, dried, and removed at reduced pressure. The yellow powder was combined and purified with silica column chromatography ethyl acetate-hexane (2:3 v/v) to yield 1 (110 mg, 19%); mp (186-188 °C).

![Scheme 7.3 Preparation of benzyl 4-(2-aminophenyl)piperazine-1-carboxylate](image)

*Reaction Conditions:* TEA and DCM.
**Experimental procedure:** 2-(piperazin-1-yl) benzenamine (0.56 mmol, 100 mg) was added to 10 mL of dichloromethane, then triethylamine (1.15 mmol, 0.16 mL) was dropped to the reaction mixture while stirring and cooled to 0 °C. Benzyl chloroformate (0.51 mmol, 0.07 ml) was added dropwise to the reaction mixture and was allowed to slowly warm to room temperature and stirring was continued overnight. The reaction was monitored by TLC detection. Reaction mixture was then diluted with more DCM, washed twice with 1M HCl and twice with brine, dried with sodium sulfate, and then concentrated at reduced pressure. The residue was purified with silica column chromatography eluting with ethyl acetate-hexane (1:4 v/v), yellow oil (116 mg, 73%).

7.6 The stereochemistry of the regioisomers

The stereochemistry of copper(II)-catalyzed C-N coupling reaction of aryl/alkyl amines to sulfonamide-substituted naphthoquinones was determined by several one-dimensional and two-dimensional NMRs, including $^1$HNMR, $^{13}$CNMR, DEPT-90, HMBC, HSQC, $^1$H$^1$HCOSY. The amination coupling was identified at position 6 but not 7, using 5,8-dioxo-6-(pyridin-3-ylamino)-5,8-dihydroronaphthalene-1-sulfonamide (3b) as an example.
Figure 7.4 $^{13}$CNMR, DEPT-90 and HMBC of compound 3b.

Figure 7.5 $^1$H$^1$HCOSY of compound 3b.
Figure 7.6 HSQC of compound 3b.
7.7 Control experiments on the reaction mechanism

Scheme 7.4 Control experiments on the reaction mechanism. **Reaction Conditions:**

(a) Cu(OAc)$_2$·H$_2$O, AcOH and H$_2$O (v/v:1/10), under N$_2$ protection, heated to 80 °C, 5h; (b) AcOH and H$_2$O (v/v:1/10), under air condition, heated to 80 °C; (c) Cu(OAc)$_2$·H$_2$O, AcOH and H$_2$O (v/v:1/10), under air condition, heated to 80 °C, overnight.

7.8 General procedure of C-N coupling reaction of ring-fused 6-amino/aniline-5H-Naphth[1,8-cd] isothiazol-5-one,1,1-dioxides

Compound 1 (237mg, 1 mmol), amine (1.2 mmol) and Cu(OAc)$_2$·H$_2$O (40 mg, 0.2 mmol) were solubilized by gently warming glacial acetic acid (5 mL) at 80-100 °C for about 3-5 hours, monitored by TLC. After 1 was completely reacted, all the volatiles were removed under reduced pressure. The resulting crude product was applied to a flash column of
silica gel, which was eluted with DCM or other solvents as indicated in the description of each product, then recrystallized in EtOH/H₂O (1:2 v/v) to give the ring-fused products (2a-2g).

7.9 Characterization data of the ring-fused compounds

6-(phenylamino)-5H-Naphth[1,8-cd]isothiazol-5-one,1,1-dioxide (2a)

The reactant amine is aniline (0.11 mL, 1.2 mmol). The product was purified by silica gel column chromatography eluting with DCM, then further purified by recrystallization. Red crystals (220 mg, yield: 71%); mp (>260 °C). ¹H NMR (300 MHz, DMSO) δ 9.78 (s, 1H), 8.39 (d, \( J = 7.1 \) Hz, 1H), 8.04 (q, \( J = 7.3 \) Hz, 2H), 7.55-7.33 (m, 4H), 7.25 (t, \( J = 6.9 \) Hz, 1H), 6.14 (s, 1H). ¹³C NMR (75 MHz, DMSO): δ = 181.38, 161.45, 144.58, 138.26, 136.84, 135.74, 129.92, 129.37, 129.03, 127.94, 126.18, 125.56, 124.44, 106.75 ppm. HRMS (ESI) of C₁₆H₁₀N₂O₃SNa [M+Na]⁺ calcd, 333.0304; found, 333.0329.

6-(pyridin-3-ylamino)-5H-Naphth[1,8-cd]isothiazol-5-one,1,1-dioxide (2b)

The reactant amine is 3-amine-pyridine (113 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with DCM, followed by DCM/EtOAc (8:1 v/v), then further purified by recrystallization. Bright red crystals (96.5 mg, yield: 31%); mp (>260 °C). ¹H NMR (300 MHz, DMSO) δ 10.39 (s, 1H), 8.53 (s, 1H), 8.36 (d, \( J = 7.9 \) Hz, 1H),
8.26 (d, J = 7.9 Hz, 1H), 8.17-8.09 (m, 2H), 8.03 (d, 1H), 7.68 (t, J = 7.5 Hz, 1H), 6.35 (s, 1H). $^{13}$C NMR (75 MHz, DMSO): $\delta = 183.63, 166.82, 147.03, 146.42, 146.10, 143.15, 135.08, 133.83, 133.75, 131.62, 131.38, 130.90, 125.02, 108.44$ ppm. HRMS (ESI) of C$_{15}$H$_9$N$_3$O$_3$SNa [M+Na]$^+$ calcd, 334.0257; found, 334.0261.

6-(1-chloro-3-nitrophenylamino)-5H-Naphth[1,8-cd]isothiazol-5-one,1,1-dioxide (2c)

The reactant amine is 2-chloro-4-nitroaniline (206 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with DCM, followed by a DCM/EtOAc (8:1 v/v), then further purified by recrystallization. Orange crystals (346 mg, yield: 89%); mp (>260 °C). $^1$H NMR (300 MHz, DMSO) $\delta$ 10.19 (s, 1H), 8.51 (d, J = 2.0 Hz, 1H), 8.45-7.93 (m, 4H), 7.83 (d, J = 8.8 Hz, 1H), 6.15 (s, 1H). $^{13}$C NMR (75 MHz, DMSO): $\delta = 177.63, 161.19, 149.92, 146.54, 141.17, 137.69, 134.80, 130.77, 129.86, 128.76, 126.97, 126.26, 124.24, 96.48$ ppm. HRMS (ESI) of C$_{16}$H$_8$ClN$_3$O$_5$SNa [M+Na]$^+$ calcd, 411.9765; found, 411.9764.

6-(naphthyl-2-amino)-5H-Naphth[1,8-cd]isothiazol-5-one,1,1-dioxide (2d)

The reactant amine is 2-naphthylamine (172 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with Hexane/DCM (1:4 v/v), followed by DCM. Red crystals (277 mg, yield: 77%), then further purified by recrystallization; mp (>260 °C). $^1$H NMR (300 MHz, DMSO) $\delta$ 9.99 (s, 1H), 8.44 (dd, J =
7.2, 1.3 Hz, 1H), 8.17-7.90 (m, 6H), 7.71-7.47 (m, 3H), 6.33 (s, 1H). $^{13}$C NMR (75 MHz, DMSO): $\delta = 181.89, 161.90, 144.80, 137.28, 136.43, 136.21, 134.17, 131.66, 130.06, 129.79, 129.43, 128.53, 128.42, 128.37, 127.66, 126.03, 123.98, 123.87, 121.73, 107.62 ppm. HRMS (ESI) of C$_{20}$H$_{12}$N$_2$O$_3$SNa $[M+Na]^+$ calcd, 383.0466; found, 383.0453.

6-(cyclohexanecarboxylic acid-3-amino)-5H-Naphth[1,8-cd]isothiazol-5-one,1,1-dioxide (2e)

The reactant amine is 3-amino-cyclohexanecarboxylic acid (172 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with DCM, followed by DCM/EtOAc (4:1 v/v). Red crystals (166 mg, yield: 46%), then further purified by recrystallization; mp (>260 °C). $^1$H NMR (300 MHz, DMSO) $\delta$ 12.16 (s, 1H), 8.59 (s, 1H), 8.27 (d, $J = 7.5$ Hz, 1H), 8.15 (d, $J = 7.6$ Hz, 1H), 7.93 (t, $J = 7.5$ Hz, 1H), 6.30 (s, 1H), 2.12-1.76 (m, 4H), 1.58-1.17 (m, 6H). $^{13}$C NMR (75 MHz, (CD$_3$)$_2$CO): $\delta = 191.20, 176.07, 160.79, 151.48, 139.91, 134.43, 129.68, 102.49, 52.47, 42.33, 31.79, 24.73, 20.98, 14.66 ppm. HRMS (ESI) of C$_{17}$H$_{16}$N$_2$O$_5$SNa $[M+Na]^+$ calcd, 383.0672; found, 383.0648.

6-(3’-phenylproanoic acid-3-amino)-5H-Naphth[1,8-cd]isothiazol-5-one,1,1-dioxide (2f)

The reactant amine is 3-((3-amino-phenyl)proanoic acid (198 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with DCM, followed by
DCM/EtOAc (4:1 v/v). Red crystals (214 mg, yield: 56%), then further purified by recrystallization; mp (>260 °C). $^1$H NMR (300 MHz, DMSO) δ 12.12 (s, 1H), 9.74 (s, 1H), 8.41 (d, $J = 7.1$ Hz, 1H), 8.16-7.96 (m, 2H), 7.45-7.08 (m, 4H), 6.16 (s, 1H), 2.87 (t, $J = 7.6$ Hz, 2H), 2.57 (t, $J = 7.3$ Hz, 2H). $^{13}$C NMR (75 MHz, DMSO): δ = 181.49, 174.18, 162.91, 150.92, 144.70, 142.84, 138.14, 136.76, 135.92, 129.92, 129.27, 127.93, 127.51, 125.46, 124.35, 112.38, 36.29, 31.42 ppm. HRMS (ESI) of C$_{19}$H$_{14}$N$_{2}$O$_{5}$SNa [M+Na]$^+$ calcd, 405.0516; found, 405.0520.

6-(benzyl carboxylate piperazine-4'-phenyl-3-amino)-5H-Naphth[1,8-cd] isothiazol-5-one,1,1-dioxide (2g)

The reactant amine is benzyl 4-(2-aminophenyl) piperazine-1-carboxylate (373 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with Hexane/DCM (1:4 v/v), followed by DCM. Red crystals (396 mg, yield: 75%), then further purified by recrystallization; mp (>260 °C). $^1$H NMR (300 MHz, DMSO) δ 9.16 (s, 1H), 8.42 (d, $J = 7.1$ Hz, 1H), 8.06 (dt, $J = 14.8$, 7.3 Hz, 2H), 7.50-7.10 (m, 9H), 5.91 (s, 1H), 5.07 (s, 2H), 3.54-3.41 (m, 4H), 2.90-2.83(m, 4H). $^{13}$C NMR (75 MHz, DMSO): δ = 181.30, 162.89, 161.84, 154.89, 154.70, 146.21, 142.52, 137.16, 136.79, 135.80, 131.29, 129.48, 128.89, 128.34, 128.01, 127.97, 127.42, 125.56, 125.04, 124.52, 121.00, 107.85, 66.80, 51.05, 36.28, 31.25 ppm. HRMS (ESI) of C$_{28}$H$_{24}$N$_{4}$O$_{5}$SNa [M + Na]$^+$ calcd, 551.1360; found, 551.1294.
7.10 General procedure of C-N coupling reaction of ring-opening 6-amino/aniline-naphthalene-5,8-dione-1-sulfonamides

Compound 1 (237mg, 1 mmol), amine (1.2 mmol) and Cu(OAc)$_2$·H$_2$O (40mg, 0.2 mmol) were solubilized by gently warming TFA and AcOH co-solvent (0.5 mL/5 mL, v/v), at 50°C to reflux, monitored by TLC. After compound 1 was completely consumed, all the volatiles were removed in vacuo to give crude products, which were purified by flash chromatography on silica gel and recrystallized in EtOH/H$_2$O (1:2 v/v) to afford the ring-opening derivatives 3a-3g.

7.11 Characterization data of the ring-opening compounds (3a-3g)

5,8-dioxo-6-(phenylamino)-5,8-dihyronaphthalene-1-sulfonamide (3a)

![Chemical structure of 3a](image)

The reactant amine is aniline (0.11 mL, 1.2 mmol). Then the product was purified by silica gel column chromatography eluting with DCM, followed by DCM/EtOAc (8:1 v/v). Red crystals (275 mg, yield: 84%); mp (>260 °C). $^1$H NMR (300 MHz, DMSO) δ 9.46 (s, 1H), 8.48-8.31 (m, 2H), 7.95(t, $J = 6.9$ Hz, 1H), 7.61-7.21 (m, 7H), 6.14 (s, 1H). $^{13}$C NMR (75 MHz, DMSO): δ = 181.73, 179.25, 144.93, 138.64, 137.20, 136.09, 130.27, 129.91, 129.71, 129.51, 128.31, 126.53, 125.92, 124.79, 104.38 ppm. HRMS (ESI) of C$_{16}$H$_{12}$N$_2$O$_4$SNa [M+Na]$^+$ calcd, 351.0415; found, 351.0423.
5,8-dioxo-6-(pyridin-3-ylamino)-5,8-dihydronaphthalene-1-sulfonamide (3b)

The reactant amine is 3-amine-pyridine (113 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with DCM/EtOAc (8:1 v/v), followed by DCM/EtOAc (2:1 v/v), then further purified by recrystallization. Bright red crystals (286 mg, yield: 87%); mp (>260 °C). \(^1\)H NMR (300 MHz, DMSO) δ 9.52 (s, 1H), 8.63 (s, 1H), 8.44 (d, \(J = 7.9\) Hz, 2H), 8.35 (d, \(J = 7.5\) Hz, 1H), 7.96 (t, \(J = 7.7\) Hz, 1H), 7.84 (d, \(J = 7.7\) Hz, 1H), 7.48 (s, 1H), 7.35 (s, 2H), 6.13 (s, 1H). \(^{13}\)C NMR (75 MHz, DMSO): δ = 183.13, 180.78, 145.92, 142.65, 134.58, 133.33, 133.25, 131.12, 130.88, 130.40, 104.56 ppm. HRMS (ESI) of C\(_{15}\)H\(_{11}\)N\(_3\)O\(_4\)SNa \([\text{M+Na}]^+\) calcd, 352.0362; found, 352.0358.

6-(1-chloro-3-nitrophenylamino)-5H-Naphth[1,8-cd]isothiazol-5-one,1,1-dioxide (3c)

The reactant amine is 2-chloro-4-nitroaniline (206 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with DCM, followed by a DCM/EtOAc (4:1 v/v), then further purified by recrystallization. Orange crystals (313 mg, yield: 77%); mp (>260 °C). \(^1\)H NMR (300 MHz, DMSO) δ 9.37 (s, 1H), 8.47 (d, \(J = 2.5\) Hz, 1H), 8.42 (d, \(J = 7.9\) Hz, 1H), 8.27 (dd, \(J = 8.9, 2.7\) Hz, 2H), 8.07 (s, 1H), 7.78 (d, \(J = 9.0\) Hz, 1H), 7.40 (s, 2H), 6.02 (s, 1H). \(^{13}\)C NMR (75 MHz, DMSO): δ = 182.16, 180.19, 169.22, 146.21, 145.21, 143.21, 141.63,
6-(naphthalen-2-ylamino)-5,8-dioxo-5,8-dihyronaphthalene-1-sulfonamide (3d)

The reactant amine is 2-naphthylamine (172 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with DCM, followed by DCM/EtOAc (8:1 v/v). Red crystals (238 mg, yield: 63%), then further purified by recrystallization; mp (>260 °C). $^1$H NMR (400 MHz, DMSO) δ 9.69 (s, 1H), 8.43-8.27 (m, 2H), 8.10-7.89 (m, 5H), 7.64-7.48 (m, 3H), 7.39 (s, 2H), 6.32 (s, 1H). $^{13}$C NMR (75 MHz, DMSO): δ = 181.99, 181.69, 147.95, 143.61, 136.49, 135.73, 134.17, 133.12, 131.56, 130.26, 129.95, 128.56, 128.48, 128.43, 127.60, 126.68, 123.90, 121.35, 102.19 ppm. HRMS (ESI) of C$_{20}$H$_{14}$N$_2$O$_4$SNa [M+Na]$^+$ calcd, 401.0566; found, 401.0576.

3-(1,4-dioxo-5-sulfamoyl-1,4-dihyronaphthalen-2-ylamino) cyclohexanecarboxylic acid (3e)

The reactant amine is 3-aminocyclohexanecarboxylic acid (172 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with DCM, followed by DCM/EtOAc (2:1 v/v). Red crystals (223 mg, yield: 59%), then further purified by recrystallization; mp (>260 °C). $^1$H NMR (300 MHz, DMSO) δ
12.10 (s, 1H), δ 8.33 (d, J = 7.1 Hz, 1H), 8.08-7.78 (m, 5H), 5.97 (s, 1H), 2.12-1.72 (m, 6H), 1.54-1.25 (m, 4H). $^{13}$C NMR (75 MHz, (CD$_3$)$_2$CO): δ = 206.51, 206.32, 206.11, 191.20, 179.08, 176.10, 150.93, 139.94, 134.46, 129.71, 127.76, 126.30, 92.58, 52.50, 42.36, 31.82, 24.77, 21.01, 14.69 ppm. HRMS (ESI) of C$_{17}$H$_{18}$N$_2$O$_6$SNa [M+Na]$^+$ calcd, 401.0778; found, 401.0796.

3-(3-(1,4-dioxo-5-sulfamoyl-1,4-dihyronaphthalen-2-ylamino)phenyl)propanoic acid (3f)

The reactant amine is 3-(3-aminophenyl)propanoic acid (198 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with DCM, followed by DCM/EtOAc (4:1 v/v). Red crystals (180 mg, yield: 45%), then further purified by recrystallization; mp (>260 ℃). $^1$H NMR (300 MHz, DMSO) δ 12.17 (s, 1H), 9.46 (s, 1H), 8.43-7.98 (m, 3H), 7.41-7.07 (m, 6H), 6.13 (s, 1H), 2.88-2.83 (t, J = 7.6 Hz, 2H), 2.59-2.54 (t, J = 7.3 Hz, 2H). $^{13}$C NMR (75 MHz, DMSO): δ = 181.88, 174.57, 143.26, 138.67, 136.14, 130.22, 128.35, 126.63, 125.68, 124.70, 122.53, 106.98, 35.82, 31.54 ppm. HRMS (ESI) of C$_{19}$H$_{16}$N$_2$O$_6$SNa [M+Na]$^+$ calcd, 423.0621; found, 423.0625.

Benzyl 4-(3-(1,4-dioxo-5-sulfamoyl-1,4-dihyronaphthalen-2-ylamino)phenyl)piperazine-1-carboxylate (3g)
The reactant amine is 4-(2-aminophenyl) piperazine-1-carboxylate (373 mg, 1.2 mmol). The product was purified by silica gel column chromatography eluting with Hexane/DCM (1:4 v/v), followed by DCM. Red crystals (186 mg, yield: 34%), then further purified by recrystallization; mp (>260 °C). $^1$H NMR (300 MHz, ) $\delta$ 8.86 (s, 1H), 8.41 (dd, $J = 26.3, 7.6$ Hz, 2H), 7.97 (t, $J = 7.9$ Hz, 1H), 7.56-7.44 (m, 1H), 7.45-7.20 (m, 10H), 6.20 (s, 1H), 5.11 (s, 2H), 3.62-3.48 (m, 4H), 2.87-2.74 (m, 4H). $^{13}$C NMR (75 MHz, DMSO): $\delta$ = 183.15, 181.02, 162.82, 157.88, 154.93, 146.28, 145.37, 143.53, 142.68, 137.22, 133.30, 131.73, 130.75, 128.91, 127.14, 126.42, 124.92, 123.21, 122.49, 104.61, 67.17, 51.45, 36.26, 31.23 ppm. HRMS (ESI) of C$_{28}$H$_{26}$O$_6$N$_4$SNa [M+Na]$^+$ calcd, 569.1465; found, 569.1460.

7.12 The hydrolysis reaction from the ring-fused compound to the ring-opening compound.

Two hydrolysis solvent systems, acids and bases, were tried. Since the PKa value of sulfonamide group is about 16.1$^{79}$, a base solution, Et$_3$N in THF, was first attempted to hydrolyze the C-N double bond. Different volume ratios of Et$_3$N and THF, from 10/90 to 100/0, were screened; however, there was no good yield of the desired product. Then, trifluoroacetic acid (TFA) was found to be able to convert 2a to 3a. A co-solvent TFA-DCM-H$_2$O (1:8:1 v/v/v) was first tried on, heated to reflux, resulting in the formation of
the desired ring-opening product 3a in 16% yield. A subsequent screening of different solvent combinations, TFA/EtOH and TFA/AcOH were found to dissolve 2a better than the initial solvent system. The yield of 3a increased to 33% in TFA/EtOH co-solvant at 40°C. Following the above results, we then performed the systematic screening on reaction temperature, the yield of 3a increased to 67% and reaction time was shortened to 5h when refluxed in TFA/EtOH co-solvant.

Scheme 7.5 Preparation of ring-opening 6-amino/anilino-naphthalene-dione-1-sulfonamide. Reaction Conditions: TFA and EtOH.

Compound 2a (237mg,1 mmol) were solubilized by gently warming EtOH and TFA co-solvent system (0.5 mL/5 mL, v/v) at 60 °C for about 3-5 hours, monitored by TLC. After 2a was completely reacted, all the volatiles were removed under reduced pressure. The resulting crude product was recrystallized in EtOH/H₂O (1:2 v/v) several times, then applied to a flash column of silica gel, which was eluted with DCM/EtOAc(8:1 v/v) then to give the ring-opening product 3a.
Chapter 8: Dynamic induced-fit docking provides insight into the STAT3 inhibition.

8.1 Abstract

To differentiate the binding affinity of structurally related compounds, an induced-fit study was performed to search for STAT3 induced structures for further drug design and development. Molecular dynamics simulations were used to compensate the requirement of the flexibility of protein structures. Systematic cross docking were applied to evaluate each protein structures and finally five induced-fit protein structures were selected for further study.

8.2 Introduction

Structure-based virtual screening (SBVS) methods have been widely utilized in drug design and discovery in the past decade\textsuperscript{2}. However, the hit rates of current structure based drug design suffer from inaccurate scoring functions and failure to incorporate protein flexibility\textsuperscript{21,80}. During the ligand binding process, the ligand and the protein adjust their conformation to achieve an overall “best-fit” and this kind of conformational adjustments (from local side chain adjustments to major domain shifts) resulting in the overall binding is referred to as “induced-fit”. In many cases, minor structural movement in a binding site
can cause major change on binding energy calculation and binding modes prediction; however, these effects are commonly ignored in rigid molecular docking and virtual screening experiments due to the high calculation cost. To rescue the false negatives and eliminate false positives, computational methods called induced-fit docking were developed to incorporate better protein flexibility, such as Schrödinger induced-fit docking (IFD) workflow. However, the accuracy of current IFD programs is still not very satisfied. To search for induced-fit structure, this study utilizes MD simulation to meet the requirement of protein flexibility. A pair of structurally similar compounds (one is active and the other is inactive) and an additional set of six known active ligands were utilized to test the induced models. Five induced-fit structures are finally selected for future drug design and discovery of STAT3 inhibitors.

8.3 Motivation of this research

In our previous structure-based drug design study, the crystal structure (PDB:1BG1) was utilized to predict functional molecules to inhibit the dimerization of pSTAT3. Two selected molecules (LY10 and LY13) were synthesized and tested in different cancer cell lines. Although LY10 and LY13 are structurally similar, they have huge difference in potency, LY13 (IC\textsubscript{50}:~1.5 μM) is much more active than LY10 (No inhibition), which piques our interest to study the induced-fit structure of the STAT3 protein.
In the docking experiments, it was hard to tell the differences in affinity between these two compounds. LY10 and LY13 were both docked to the STAT3 crystal structure (PDB: 1BG1). The energy of LY10 is -9.3 Kcal/mol and the energy of LY13 is -9.0 Kcal/mol by Autodock4. To explain the difference in biological activity, one hypothesis is the rigid crystal structure of STAT3 might not be the realistic structure when ligand approaches to the STAT3 protein, which results in the failure of the crystal structure to differentiate the false positive. So the aim of this study is to search for induced-fit structures which can simulate the realistic structure and can be used in virtual screening to minimize the error rates of the false positive and false negative inhibitors.

8.4 Materials and Methods

**STAT3 protein**: PDB: 1BG1; **Known active inhibitors**: LY13, LLL12, Cpd188, S31-M2001, S31-201, PpYL, Wyeth. **Known inactive decoy probe**: LY10
8.5 Conformational sampling of apo-STAT3 and holo-STAT3-ligand complex

An ensemble of apo and holo-STAT3 snapshots was collected from a molecular dynamics (MD) trajectory to account for protein flexibility during docking. Constant temperature and pressure MD simulation of the STAT3 models was performed by Amber using the SHAKE algorithm. Since we are interested in a small scale conformational change of the binding sites in SH2 domain of STAT3, to save the computational cost, residues (Gly499-Arg688) were saved for the simulation study. The MD system was set up as follow. Step 1: the STAT3 protein system was solvated with a 15 Å thick truncated octahedron TIP3P water model, neutralized by Na⁺ ions, and energy minimized by 100
steps of steepest descent minimization followed by 2000 steps of conjugate gradient minimization. The system was equilibrated by a heating up system from 100k to 300k over 25 ps. Then a 0.1 ns MD with constant temperature and pressure in the explicit water model completed the equilibration stage. Finally, an implicit water topology file was created for the MD run. Step 5: Trajectories in ~5 ns used a heat bath coupling constant of 2.0 ps and were made at 1 atm with a pressure relaxation performed on the IBM Opteron cluster using 4 cores (4 processors per MD). All simulations were run using the sander code of Amber 9 with the ff99 force field.

8.6 Extract representative structures of each MD run

Trajectories of each MD are clustered into five distinct sets using the most common similarity metric root-mean-square deviation (RMSD) algorithms. The detailed procedures are as follow. Step 1: Trajectory files were read in and all the snapshots were aligned to the first snapshot based on all the heavy atoms. New trajectory files were saved after water and ions were depleted. Step 2: Each MD simulation used the average-linkage algorithm to produce five clusters using the pairwise RMSD between frames as a metric comparing the atoms named CA. Five representative structures were saved as PDB format from the five clusters of each MD run. In total, 35 representative PDBs from one apo and six holo MD runs were prepared for the further screening.
8.7 Ensemble Dockings

LY10 was used as a decoy probe and was expected not to recognize the ligand-induced fit structure of the STAT3 dimerization interface. For the fast dockings to the large set of STAT3 representative structures (5 apo and 30 holo structures), Glide in the Schrödinger package was used to perform cross-docking.

All the representative structures were subjected to preliminary preparation for docking employing the Schrödinger Protein Preparation Wizard protocol of Schrödinger Suite 2010. Hydrogen atoms were added followed by proper protonation state assignment. The structures were minimized with cut off criteria of 0.30 Å RMSD using the OPLS2005 force field. Docking grids were constructed for each snapshot by Receptor Grid Generation. Grid boxes capable of accommodating ligands with up to 20 Å of geometric length were created with the centroid on residue Arg609. Both active and inactive compounds were docked to the prepared grids by Glide with docking precision XP (extra precision), and then Gscores were analyzed.

8.8 Structural analysis of the induced-fit structures of STAT3

The five selected induced-fit structures were aligned to the crystal structure by Schrödinger Protein Structure Alignment tools. The residue-wise RMSD was calculated by the Schrödinger script “rmsd_by_residue.py” and plotted by R program.
8.9 Results

**Docking experiment of known active compounds to the crystallized structure of STAT3**

All the known STAT3 inhibitors and known inactive decoy compound LY10 were docked to the crystal structure of STAT3 (PDB: 1BG1). From the result as shown in Table 8.1, LY10 was found as a false positive from the docking experiment compared to the tested IC\textsubscript{50} values of the known STAT3 inhibitors, which indicates that the crystal structure may not be the real structure of STAT3 during the realistic binding process. So the search for an induced-fit structure of STAT3 is necessary.

**Conformational sampling of apo-STAT3 and holo-STAT3-ligand complex**

In order to identify the induced-fit structures for further drug design, a systematic conformational sampling were performed by MD simulation and an ensemble of representative apo-STAT3 and holo-STAT3 snapshots were clustered out to account for protein flexibility during docking. The RMSD of the 5ns MD simulations of STAT3 protein with/without known active inhibitors (LLL12, Cpd188, S31-M2001, S31-201, PpYL and Wyeth) are shown in Figure 8.2.
Table 8.1 The comparison of docking scores and IC$_{50}$ values of the known STAT3 inhibitors and decoy compound.

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Gscore$^a$ (Kcal/mol)</th>
<th>IC$_{50}$(μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LY10</td>
<td>-5.5</td>
<td>No inhibition (MTT cell viability assay)</td>
</tr>
<tr>
<td>S3I-M2001</td>
<td>-5.4</td>
<td>79 (WST-1 cell viability assay)</td>
</tr>
<tr>
<td>PpYL</td>
<td>-5.4</td>
<td>182 (DB$_{50}^b$)</td>
</tr>
<tr>
<td>LY13</td>
<td>-5.4</td>
<td>1.5 (MTT cell viability assay)</td>
</tr>
<tr>
<td>Wyeth</td>
<td>-4.5</td>
<td>357 (DB$_{50}$)</td>
</tr>
<tr>
<td>Cpd188</td>
<td>-4.5</td>
<td>73 (IC$_{50}$ by inhibition of IL-6-mediated phosphorylation of Stat3)</td>
</tr>
<tr>
<td>S3I-201</td>
<td>-4.2</td>
<td>86 (DB$_{50}$)</td>
</tr>
<tr>
<td>LLL-12</td>
<td>-1.9</td>
<td>0.16-3.09 (MTT cell viability assay)</td>
</tr>
</tbody>
</table>

$^a$Gscores were calculated by Schrödinger Glide with docking precision XP (extra precision). $^b$DB$_{50}$, concentration of peptide at which DNA binding activity is reduced by 50%.
Figure 8.2 The RMSDs of the MD simulations of apo-STAT3 and holo-STAT3.

Extract representative structures of each MD run

All the snapshots of each MD were clustered into five groups using the pairwise RMSD method, so there were total 35 clusters for 7 MD runs and 35 representative structures were saved from each cluster, named as apo-STAT3.C0 to apo-STAT3.C4 and STAT3-ligand.C0 to STAT3-ligand.C4.

Ensemble Dockings and energy analysis of the induced structures

Eight compounds (LLL12, Cpd188, S31-M2001, S31-201, PpYL, Wyeth, LY13 and decoy LY10) were docked to the 35 representative structures by Glide with docking
precision XP (extra precision), and then Gscores were recorded (see supporting information) and analyzed.

Instead of identify an induced-fit structure based on the statistical frequency or the enrichment rates of the frames\textsuperscript{18,81}, a tested structurally similar decoy compound can provide more useful information which helps to screen out the inaccurate structures.

In virtual screening, the docking score is set as one of the most crucial criteria for the selection of hits. The docking scores of all known inhibitors were subtracted by the Gscores of LY10 docked to all the structures. All the structures were ranked by the sum of the $\Delta G_{\text{ligand-LY10}}$ (Table 8.2). Five induced-fit structures, apo-STAT3.C3, holo-STAT3-S31-201.C1, holo-STAT3-Wyeth.C4, holo-STAT3-LLL12.C2 and holo-STAT3-LLL12.C3, were selected, which successfully differentiate the false positive compound LY10 based on energy analysis. And the $\Delta G_{\text{ligand-LY10}}$ surface provides us a general view of all the snapshots (Figure 8.3).
Figure 8.3 The $\Delta G_{\text{ligand-LY10}}$ surfaces of all the snapshots.
Table 8.2 The Gscore difference between active compounds and LY10.

<table>
<thead>
<tr>
<th></th>
<th>ΔG_{ligand-LY10}(Kcal/mol)</th>
<th>LY13</th>
<th>LLL12</th>
<th>Cpd188</th>
<th>S31-M2001</th>
<th>S31-201</th>
<th>PpYL</th>
<th>Wyeth</th>
<th>Sum of ΔG_{ligand-LY10}</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT3-LLL12.c2</td>
<td>-3.5</td>
<td>-4.2</td>
<td>-3.0</td>
<td>-1.6</td>
<td>-2.3</td>
<td>-6.0</td>
<td>-3.4</td>
<td>-24.1</td>
<td></td>
</tr>
<tr>
<td>STAT3-S31-201.c1</td>
<td>-3.0</td>
<td>-1.0</td>
<td>-3.7</td>
<td>-1.8</td>
<td>-2.5</td>
<td>-3.1</td>
<td>-1.1</td>
<td>-16.4</td>
<td></td>
</tr>
<tr>
<td>STAT3-Cpd188.c3</td>
<td>-5.2</td>
<td>-0.9</td>
<td>-1.1</td>
<td>-2.2</td>
<td>0.3</td>
<td>-5.1</td>
<td>-1.7</td>
<td>-15.7</td>
<td></td>
</tr>
<tr>
<td>STAT3-LLL12.c3</td>
<td>-2.0</td>
<td>-1.3</td>
<td>-1.4</td>
<td>-0.3</td>
<td>-3.2</td>
<td>-5.1</td>
<td>-2.7</td>
<td>-15.9</td>
<td></td>
</tr>
<tr>
<td>apo-STAT3.c0</td>
<td>-3.2</td>
<td>0.0</td>
<td>-2.0</td>
<td>-1.3</td>
<td>-1.0</td>
<td>-3.9</td>
<td>-3.1</td>
<td>-14.5</td>
<td></td>
</tr>
<tr>
<td>STAT3-LLL12.c0</td>
<td>-3.2</td>
<td>0.5</td>
<td>-1.3</td>
<td>0.0</td>
<td>-1.5</td>
<td>-3.9</td>
<td>-1.9</td>
<td>-11.3</td>
<td></td>
</tr>
<tr>
<td>apo-STAT3.c3</td>
<td>-0.6</td>
<td>1.9</td>
<td>-1.4</td>
<td>-1.4</td>
<td>-0.9</td>
<td>-4.2</td>
<td>-2.0</td>
<td>-12.3</td>
<td></td>
</tr>
<tr>
<td>STAT3-Wyeth.c0</td>
<td>-1.2</td>
<td>-2.4</td>
<td>-2.6</td>
<td>-0.2</td>
<td>-4.3</td>
<td>-3.0</td>
<td>1.2</td>
<td>-12.4</td>
<td></td>
</tr>
<tr>
<td>apo-STAT3.c4</td>
<td>-2.6</td>
<td>-0.6</td>
<td>-0.8</td>
<td>0.4</td>
<td>-0.6</td>
<td>-3.3</td>
<td>-1.8</td>
<td>-9.4</td>
<td></td>
</tr>
<tr>
<td>STAT3-LLL12.c1</td>
<td>-0.9</td>
<td>-0.8</td>
<td>-0.1</td>
<td>0.9</td>
<td>-3.3</td>
<td>-3.5</td>
<td>-1.4</td>
<td>-9.2</td>
<td></td>
</tr>
<tr>
<td>STAT3-Wyeth.c4</td>
<td>-2.1</td>
<td>1.4</td>
<td>-1.1</td>
<td>-2.0</td>
<td>-0.9</td>
<td>-2.0</td>
<td>-0.3</td>
<td>-9.6</td>
<td></td>
</tr>
<tr>
<td>STAT3-Cpd188.c0</td>
<td>-1.7</td>
<td>0.6</td>
<td>-3.0</td>
<td>0.5</td>
<td>-0.7</td>
<td>-3.0</td>
<td>-0.1</td>
<td>-8.7</td>
<td></td>
</tr>
<tr>
<td>STAT3-PpYL.c3</td>
<td>-1.0</td>
<td>2.0</td>
<td>-1.8</td>
<td>-0.3</td>
<td>-2.4</td>
<td>-2.5</td>
<td>0.7</td>
<td>-5.2</td>
<td></td>
</tr>
<tr>
<td>STAT3-LLL12.c4</td>
<td>-2.7</td>
<td>0.2</td>
<td>-0.7</td>
<td>0.9</td>
<td>0.9</td>
<td>-4.8</td>
<td>0.7</td>
<td>-6.8</td>
<td></td>
</tr>
<tr>
<td>STAT3-S31-201.c3</td>
<td>-1.7</td>
<td>-1.0</td>
<td>-0.3</td>
<td>0.2</td>
<td>-0.2</td>
<td>-2.4</td>
<td>-2.4</td>
<td>-8.0</td>
<td></td>
</tr>
<tr>
<td>STAT3-S31-201.c0</td>
<td>-2.3</td>
<td>0.0</td>
<td>-1.5</td>
<td>0.4</td>
<td>0.0</td>
<td>-2.7</td>
<td>-0.3</td>
<td>-6.3</td>
<td></td>
</tr>
<tr>
<td>apo-STAT3.c1</td>
<td>-1.7</td>
<td>0.3</td>
<td>0.6</td>
<td>0.9</td>
<td>-1.1</td>
<td>-4.3</td>
<td>0.3</td>
<td>-5.0</td>
<td></td>
</tr>
<tr>
<td>STAT3-S31-M2001.c2</td>
<td>-3.0</td>
<td>4.8</td>
<td>-3.0</td>
<td>2.3</td>
<td>-4.1</td>
<td>-1.7</td>
<td>4.5</td>
<td>-0.2</td>
<td></td>
</tr>
<tr>
<td>STAT3-PpYL.c0</td>
<td>-0.9</td>
<td>1.6</td>
<td>-0.9</td>
<td>0.1</td>
<td>-2.1</td>
<td>-2.2</td>
<td>1.0</td>
<td>-3.3</td>
<td></td>
</tr>
<tr>
<td>STAT3-S31-201.c4</td>
<td>-2.3</td>
<td>1.1</td>
<td>0.6</td>
<td>0.3</td>
<td>0.6</td>
<td>-1.4</td>
<td>-0.4</td>
<td>-2.7</td>
<td></td>
</tr>
<tr>
<td>STAT3-Wyeth.c3</td>
<td>-0.4</td>
<td>2.0</td>
<td>-1.3</td>
<td>1.1</td>
<td>-0.4</td>
<td>-2.1</td>
<td>-0.3</td>
<td>-1.3</td>
<td></td>
</tr>
<tr>
<td>STAT3-S31-M2001.c1</td>
<td>-0.9</td>
<td>2.1</td>
<td>-0.3</td>
<td>0.2</td>
<td>-1.2</td>
<td>-3.0</td>
<td>1.9</td>
<td>-1.2</td>
<td></td>
</tr>
<tr>
<td>apo-STAT3.c2</td>
<td>-2.1</td>
<td>0.1</td>
<td>-1.0</td>
<td>2.6</td>
<td>0.0</td>
<td>-2.8</td>
<td>0.1</td>
<td>-3.2</td>
<td></td>
</tr>
<tr>
<td>STAT3-Wyeth.c1</td>
<td>-1.1</td>
<td>1.5</td>
<td>-1.0</td>
<td>0.7</td>
<td>-0.2</td>
<td>-2.0</td>
<td>0.8</td>
<td>-1.4</td>
<td></td>
</tr>
<tr>
<td>STAT3-S31-M2001.c0</td>
<td>-1.3</td>
<td>3.8</td>
<td>0.0</td>
<td>-0.9</td>
<td>-0.1</td>
<td>-2.6</td>
<td>3.1</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>STAT3-PpYL.c1</td>
<td>-1.2</td>
<td>3.0</td>
<td>-0.3</td>
<td>0.6</td>
<td>-1.3</td>
<td>-0.4</td>
<td>1.2</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>STAT3-PpYL.c4</td>
<td>-0.4</td>
<td>2.3</td>
<td>-1.1</td>
<td>-0.3</td>
<td>1.0</td>
<td>-3.2</td>
<td>2.8</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>STAT3-PpYL.c2</td>
<td>-0.6</td>
<td>2.6</td>
<td>-0.4</td>
<td>0.3</td>
<td>-0.2</td>
<td>-0.8</td>
<td>2.9</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>STAT3-Wyeth.c2</td>
<td>-2.0</td>
<td>1.3</td>
<td>1.1</td>
<td>2.1</td>
<td>0.3</td>
<td>-1.6</td>
<td>1.9</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>STAT3-Cpd188.c1</td>
<td>-2.0</td>
<td>4.0</td>
<td>-0.1</td>
<td>2.8</td>
<td>-1.1</td>
<td>0.0</td>
<td>3.2</td>
<td>6.8</td>
<td></td>
</tr>
<tr>
<td>STAT3-Cpd188.c4</td>
<td>-1.8</td>
<td>1.9</td>
<td>0.4</td>
<td>3.0</td>
<td>2.4</td>
<td>-1.9</td>
<td>1.2</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>STAT3-S31-201.c2</td>
<td>0.6</td>
<td>1.6</td>
<td>0.5</td>
<td>2.8</td>
<td>-0.2</td>
<td>0.5</td>
<td>1.7</td>
<td>7.6</td>
<td></td>
</tr>
<tr>
<td>STAT3-S31-M2001.c4</td>
<td>-2.2</td>
<td>5.5</td>
<td>1.6</td>
<td>3.7</td>
<td>-1.2</td>
<td>-0.1</td>
<td>5.0</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>STAT3-S31-M2001.c3</td>
<td>-1.1</td>
<td>5.0</td>
<td>1.4</td>
<td>2.3</td>
<td>-0.6</td>
<td>0.6</td>
<td>4.8</td>
<td>12.3</td>
<td></td>
</tr>
<tr>
<td>STAT3-Cpd188.c2</td>
<td>1.7</td>
<td>5.8</td>
<td>1.9</td>
<td>5.1</td>
<td>4.3</td>
<td>0.1</td>
<td>4.0</td>
<td>23.0</td>
<td></td>
</tr>
</tbody>
</table>

*The Gscore energy difference above 0 were colored as pink.*

116
Structural analysis of the induced-fit structures of STAT3

The five selected induced-fit structures were further compared with the crystal structure of STAT3. The residuewise RMSD plot (Figure 8.4) shows that the two most fluctuant regions of protein are residues 622-627 and residues 661-665 (Figure 8.5).

Figure 8.4 The comparison of residue-wise RMSDs of the selected induced-fit structures and the crystal structure (PDB:1BG1).
Figure 8.5 Two most fluctuant regions of protein, residues 622-627 and residues 661-665, are in two neighbor loops.

In order to investigate the alternation in binding sites, the surfaces of five induced-fit structures are compared to the molecular surface of the crystal structure of STAT3. The two original binding sites, side pocket and the pTyr705 site, are connected to the lower hydrophobic pocket and the site pocket binding site becomes shallower than the original site, which helps to differentiate the LY10 and LY13 (Figure 8.6).
Figure 8.6 A parallel comparison of the crystal structure and the five selected induced-fit structures.
8.10 Discussion

To explain the significant difference in potency of LY10 and LY13, seven MD runs were performed to provide enough protein flexibility. Seven known active compounds and one inactive compound were used to select the induced-fit structures based on the sum of the $\Delta G_{\text{ligand-LY10}}$. From the comparison of the induced-fit structures and the crystal structure, some major modifications were discovered, especially the site pocket was greatly shrunk in these induced-fit structures, but the pTYR705 site was extended and connected to the lower portion of the protein. LY10 with a long side chain, phenylpropionic acid, in its structure might cause steric hindrances with the shrunk side pocket. In future drug design, the fragment binding to the side pocket should not be too large. And a fragment which binds to the lower portion of pTYR705 site might increase the whole molecule’s binding affinity. In addition, the five selected structures can be used as a last step filtration in virtual screening to screen out the false positives too.

8.11 Conclusion

In this study, five induced-fit structures were screened out for further virtual screening and structure-based drug design. The protein flexibility was prepared by seven MD simulations. Snapshots were clustered by a RMSD based algorithm. A reliable compound
set with seven known active compounds and one inactive compound was used to finally filter out the induced-fit structures, which can be applied in future study. Based on the structural analysis, the alternation of the side pocket also indicates that the size of fragment for side pocket should not be too large in order to avoid the potential steric crash with the STAT3 protein.
Chapter 9: Summary and Conclusions

In summary, this research firstly developed a novel *in silico* site-directed fragment-based drug design method for recombination of multiple fragments from existing inhibitors. As an efficient and feasible method, it was applied in the design and discovery of novel potent STAT3 inhibitors. Several potent STAT3 inhibitors were discovered represented by LY5 (Chapter 2). Further SAR study was performed, the cyclized compounds were found to be generally more active than the ring-opening compounds. And the potency of the cyclized compound LY5-1 was found to be improved or at least comparable to the activity of LY5 in two cell lines (Chapter 3). Utilizing another improved fragment-based drug design method, multiple ligand simultaneous docking, several new STAT3 inhibitors were further discovered with better predicted bioavailability, represented by LY17 with a $K_D$ value about 2.4 μM (Chapter 4). In addition, a highly regioselective one-pot C-N coupling reaction and chemoselective ring-fusion/ring-opening synthetic methods were discovered. This protocol can not only serve as a new avenue for amination coupling of sensitive functional groups substituted naphthoquinone, but also afford a facile approach for the synthesis of various bioactive molecules (Chapter 5). Finally, the induced-fit STAT3 proteins structures were investigated for further drug design and development (Chapter 6). Overall, this study used many kinds of techniques and knowledge in computational drug design, organic synthesis, biochemical assays and biological studies to discover novel STAT3 inhibitors for anticancer therapy.
References


Hsu, Y. L., *et al.* Plumbagin (5-hydroxy-2-methyl-1,4-naphthoquinone) induces apoptosis and cell cycle arrest in A549 cells through p53 accumulation via c-Jun NH2-


Appendix A. Copies $^1$H NMR, $^{13}$C NMR.

6-(phenylamino)-5$H$-naphth[1,8-$cd$]isothiazol-5-one, 1,1-dioxide (2a)
6-(pyridin-3-ylamino)-5H-naphth[1,8-cd]isothiazol-5-one,1,1-dioxide (2b)
6-(1-chloro-3-nitrophenylamino)-5H-naph[1,8-cd]isothiazol-5-one,1,1-dioxide (2c)
6-(naphthyl-2-amino)-5H-naphth[1,8-cd]isothiazol-5-one,1,1-dioxide (2d)
6-(cyclohexanecarboxylic acid-3-amino)-5H-naphth[1,8-\textit{cd}]isothiazol-5-one,1,1-dioxide (2e)
6-(3’-phenylpropanoic acid-3-amino)-5H-naphth[1,8-cd]isothiazol-5-one, 1,1-dioxide (2f)
6-(benzyl carboxylate piperazine-4'-phenyl-3-amino)-5H-naphth[1,8-cd] isothiazol-5-one,1,1-dioxide (2g)
5,8-dioxo-6-(phenylamino)-5,8-dihyronaphthalene-1-sulfonamide (3a)
5,8-dioxo-6-(pyridin-3-ylamino)-5,8-dihydronaphthalene-1-sulfonamide (3b)
6-(1-chloro-3-nitrophenylamino)-5H-Naphth[1,8-cd]isothiazol-5-one,1,1-dioxide (3c)
6-(naphthalen-2-ylamino)-5,8-dioxo-5,8-dihydronaphthalene-1-sulfonamide (3d)
3-(1,4-dioxo-5-sulfamoyl-1,4-dihydronaphthalen-2-ylamino) cyclohexanecarboxylic acid (3e)
3-(3-(1,4-dioxo-5-sulfamoyl-1,4-dihydrornaphthalen-2-ylamino)phenyl)propanoic acid (3f)
Benzyl 4-(3-(1,4-dioxo-5-sulfamoyl-1,4-dihydronaphthalen-2-ylamino)phenyl) piperazine-1-carboxylate (3g)
Appendix B. The docked fragment combinations by MLSD.

Figure B.1 Fragment 1 and polar-aromatic-amine fragments.
Figure B.2 Fragment 1 and polar-aromatic-amine fragments with hydroxyl substitution.

Figure B.3 Fragment 1 and polar-aromatic-sulfonamide/polar-aromatic-cyclic amine fragments.
Figure B.4 Fragment 1 and polar-aromatic acid fragments.
Appendix C. The reposition of selected compounds.

Figure C.1 The reposition of LY9 to the docked fragments.

Figure C.2 The reposition of LY10 to the docked fragments.
Figure C.3 The reposition of LY11 to the docked fragments.

Figure C.4 The reposition of LY17 to the docked fragments.
Appendix D Computational drug design of natural products as STAT3 inhibitors

There are many known natural products reported as STAT3 inhibitors, such as Curcumin, Cucurbitacin Q.

From the docking experiments, we predicted that the following listed natural compounds may have potential activity on STAT3 SH2 domain.

Figure D.1 Sesquiterpene lactone and its docking mode to STAT3 SH2 domain.
Figure D.2 Sesquiterpene lactone derivative and its docking mode to STAT3 SH2 domain.

Figure D.3 4-Beta-Hydroxywithanolide E and its docking mode to STAT3 SH2 domain.