

Anthony E. Klon *Editor*

# Fragment-Based Methods in Drug Discovery



Humana Press

# METHODS IN MOLECULAR BIOLOGY

*Series Editor*

John M. Walker

School of Life and Medical Sciences

University of Hertfordshire

Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:  
<http://www.springer.com/series/7651>



# **Fragment-Based Methods in Drug Discovery**

Edited by

**Anthony E. Klon**

*Pennsylvania Drug Discovery Institute, Doylestown, PA, USA*



*Editor*

Anthony E. Klon  
Pennsylvania Drug Discovery Institute  
Doylestown, PA, USA

ISSN 1064-3745

Methods in Molecular Biology

ISBN 978-1-4939-2485-1

DOI 10.1007/978-1-4939-2486-8

ISSN 1940-6029 (electronic)

ISBN 978-1-4939-2486-8 (eBook)

Library of Congress Control Number: 2015931254

Springer New York Heidelberg Dordrecht London  
© Springer Science+Business Media New York 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Humana Press is a brand of Springer

Springer Science+Business Media LLC New York is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

---

## Preface

Fragment-based methods for drug discovery have been investigated in one form or another for several decades, but there has been increased interest in the last 10 years in their practical application in drug discovery. This is partly due to some of the recent successes of the field and their contribution to drug discovery, as well as an expansion in the number and availability of methods and improved computational resources. This volume will cover the techniques necessary for a successful fragment-based drug design project, beginning from defining the problem in terms of preparing the protein model, identifying potential binding sites, and the consideration of various candidate fragments for simulation. The second part of this volume discusses the technical aspects that various methods have used to simulate fragment binding to a target protein using Monte Carlo, molecular dynamics, and docking algorithms. After simulations, fragments are assembled into molecules using a variety of approaches, which are explored next. A discussion of design strategies and consideration of drug-like properties is included as part of the design process at this stage. Finally, several examples of successful fragment-based drug design projects are presented.

*Doylestown, PA*

*Anthony E. Klon*



---

# Contents

Preface .....	v
Contributors .....	ix

## PART I PREPARATION

1 Solvation Methods for Protein–Ligand Docking .....	3
<i>Rachelle J. Bienstock</i>	
2 Binding Site Druggability Assessment in Fragment-Based Drug Design.....	13
<i>Tu Zhou and Niu Huang</i>	
3 Generating “Fragment-Based Virtual Library” Using Pocket Similarity Search of Ligand–Receptor Complexes .....	23
<i>Raed S. Khashan</i>	
4 Virtual Fragment Preparation for Computational Fragment-Based Drug Design.....	31
<i>Jennifer L. Ludington</i>	
5 Fragment Library Design: Using Cheminformatics and Expert Chemists to Fill Gaps in Existing Fragment Libraries .....	43
<i>Peter S. Kutchukian, Sung-Sau So, Christian Fischer, and Chris L. Waller</i>	

## PART II SIMULATION

6 Protocol for Fragment Hopping .....	57
<i>Kevin B. Teuscher and Haitao Ji</i>	
7 Site Identification by Ligand Competitive Saturation (SILCS) Simulations for Fragment-Based Drug Design.....	75
<i>Christina E. Faller, E. Prabhu Raman, Alexander D. MacKerell, Jr., and Olgun Guvench</i>	
8 A Computational Fragment-Based De Novo Design Protocol Guided by Ligand Efficiency Indices (LEI) .....	89
<i>Álvaro Cortés-Cabrera, Federico Gago, and Antonio Morreale</i>	
9 Scoring Functions for Fragment-Based Drug Discovery.....	101
<i>Jui-Chih Wang and Jung-Hsin Lin</i>	

## PART III DESIGN

10 Computational Methods for Fragment-Based Ligand Design: Growing and Linking .....	119
<i>Rachelle J. Bienstock</i>	
11 Design Strategies for Computational Fragment-Based Drug Design .....	137
<i>Zenon D. Konteatis</i>	

12	Protein Binding Site Analysis for Drug Discovery Using a Computational Fragment-Based Method. . . . .	145
	<i>Jennifer L. Ludington</i>	
<b>PART IV CASE STUDIES</b>		
13	Fragment-Based Design of Kinase Inhibitors: A Practical Guide. . . . .	157
	<i>Jon A. Erickson</i>	
14	Designing a Small Molecule Erythropoietin Mimetic. . . . .	185
	<i>Frank Guarnieri</i>	
15	Designing an Orally Available Nontoxic p38 Inhibitor with a Fragment-Based Strategy. . . . .	211
	<i>Frank Guarnieri</i>	
	<i>Index</i> . . . . .	227

---

## Contributors

RACHELLE J. BIENSTOCK • *Independent Researcher and Consultant, Chapel Hill, NC, USA*

ÁLVARO CORTÉS-CABRERA • *Unidad de Bioinformática, Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Madrid, Spain; Área de Farmacología, Departamento de Ciencias Biomédicas, Universidad de Alcalá, Madrid, Spain*

JON A. ERICKSON • *Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN, USA*  
CHRISTINA E. FALLER • *Department of Pharmaceutical Sciences, College of Pharmacy, University of New England, Portland, ME, USA*

CHRISTIAN FISCHER • *Discovery Chemistry, Merck Research Laboratories, Boston, MA, USA*  
FEDERICO GAGO • *Área de Farmacología, Departamento de Ciencias Biomédicas, Universidad de Alcalá, Madrid, Spain*

FRANK GUARNIERI • *Department of Physiology & Biophysics, Virginia Commonwealth University School of Medicine, Richmond, VA, USA; Phase III Pharmaceuticals, Brooklyn, NY, USA*

OLGUN GUVENCH • *Department of Pharmaceutical Sciences, College of Pharmacy, University of New England, Portland, ME, USA*

NIU HUANG • *Dr. Niu Huang's Lab, National Institute of Biological Sciences, Beijing, China*  
HAITAO JI • *Department of Chemistry, Center for Cell and Genome Science, University of Utah, Salt Lake City, UT, USA*

RAED S. KHASHAN • *Department of Pharmaceutical Sciences, College of Clinical Pharmacy, King Faisal University, Al-Ahsa, Saudi Arabia; Laboratory for Molecular Modeling, Division of Chemical Biology & Medicinal Chemistry, The UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*

ZENON D. KONTEATIS • *Agios Pharmaceuticals, Cambridge, MA, USA*

PETER S. KUTCHUKIAN • *Merck Research Laboratories, Boston, MA, USA*

JUNG-HSIN LIN • *Division of Mechanics, Research Center for Applied Sciences, and Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; School of Pharmacy, National Taiwan University, Taipei, Taiwan*

JENNIFER L. LUDINGTON • *Locus Pharmaceuticals, Inc., Blue Bell, PA, USA*

ALEXANDER D. MACKERELL, JR. • *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA*

ANTONIO MORREALE • *Unidad de Bioinformática, Centro de Biología Molecular Severo Ochoa, Madrid, Spain; Repsol Technology Center, Madrid, Spain*

E. PRABHU RAMAN • *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA*

SUNG-SAU SO • *Computational Chemistry, Merck Research Laboratories, Kenilworth, NJ, USA*

KEVIN B. TEUSCHER • *Department of Chemistry, Center for Cell and Genome Science, University of Utah, Salt Lake City, UT, USA*

CHRIS L. WALLER • *Scientific Modeling Platforms, Merck Research Laboratories, Boston, MA, USA*

JUI-CHIH WANG • *Division of Mechanics, Research Center for Applied Sciences, Academia Sinica, Taipei, Taiwan*

YU ZHOU • *Dr. Niu Huang's Lab, National Institute of Biological Sciences, Beijing, China*

# **Part I**

## **Preparation**

# Chapter 1

## Solvation Methods for Protein–Ligand Docking

Rachelle J. Bienstock

### Abstract

Hydration has a significant impact on ligand binding within protein active sites. Specific water molecules and their placement within protein active sites have been shown to make specific contributions to the energetics of protein–ligand binding and need consideration in the design of efficient binding ligands. These specific nonbulk water molecules and their interactions are different and have more significant impact in ligand design than the generalized bulk solvation of ligand–protein systems. Proper theoretical description of the solvation effects of water within a ligand-binding pocket is a significant computational challenge. Recently, new computational methods have been developed which can more accurately describe the contribution of waters within a protein ligand site and lead to improved and enhanced ligand design and ranking in computational docking and to greater enrichment.

**Key words** Solvation models, 3D-RISM, Watermap, Explicit solvation, Implicit solvation

### Abbreviations

RISM	Reference interaction site model
IFST/IST	Inhomogeneous fluid solvation theory
HFE	Hydration free energy
MM-PB/SA	Molecular mechanics-poisson-boltzmann/surface area
MM-GB/SA	Molecular mechanics-generalized born method/surface area
FEP	Free energy perturbation
TIP	Transferable interaction potential
MD	Molecular dynamics
IETs	Integral equation theories

---

### 1 Introduction

There is increasing interest in understanding the role and energy contribution of water molecules within protein active sites in drug design, since these nonbulk waters have been shown to significantly impact protein–ligand binding. Ligand affinity and

specificity has been linked to and affected by water molecule interactions within ligand–receptor complexes. When a ligand binds, in a protein-binding pocket, solvent molecules within the pocket rearrange or become displaced. These rearrangements of water molecules within the binding pocket affect the binding free energy. Water solvation of narrow hydrophobic pockets is very unfavorable energetically because water in this situation cannot form hydrogen bonds. Expulsion of water from these types of enclosed protein regions leads to enhancement in ligand–protein binding affinity. Displacement of water molecules, from the active site, when a ligand that is more energetically suitable binds, can liberate free energy when the ligand binding displaces active site waters (HFE=hydration free energy). Free energy perturbation methods can compute this free energy gain but at a significant computational price. Are there less computationally expensive methods that can be used to calculate these free energy perturbation gains and therefore give a more accurate docking score and measure of ligand-binding affinity?

There are computational methods for including hydration effects ranging from simplistic empirical approximations to full rigorous free energy perturbation theoretical treatments. Can continuum solvation theories, which have been the major computational solvation methods used to describe hydration up until the present, adequately describe the physics involved to rank binding affinities of ligands? Which methods are the most suitable and which can be used to effectively calculate ligand binding energies and correctly rank docked ligands?

Methods have been developed to calculate the contribution of the solvent to the binding free energy of a small molecule to its receptor, that include the effects of the ligand displacing solvent from the protein active site. These methods fall into three general categories: (Subheading 1.1) Explicit Solvation Simulation, (Subheading 1.2) Implicit Solvation (Solvent Continuum), i.e., includes only the high dielectric polarization term for water, and (Subheading 1.3) Intermediate Approximation Methods, i.e., Solvent Potential Methods.

### **1.1 Explicit Solvation—Explicit Water Models**

Explicit water models differ in geometry, electrostatics, and parameters. TIP3P (3-point transferable interaction potential) water [1] and SPC water (simple point charge) [2] are among the most common water models used. They are fixed charge, rigid water models. Efforts to improve the three-site TIP model have resulted in TIP4P, TIP5P water models [1]. Explicit solvent models use a microscopic and atomistic description with each water molecule represented as a point charge with Lennard-Jones van der Waals potential representations. The advantage of explicit water models is that they can investigate reaction dynamics and probe the specific structure of water. However, explicitly treating each water molecule is more

computationally intensive. Explicit methods to calculate the binding free energy of displaced waters, and whether water displacement is energetically favorable, include Thermodynamic integration (TI) or free energy perturbation theory (FEP) [3], which are used to calculate the energetics of water molecules in the binding pocket.

### 1.1.1 IFST (Inhomogeneous Fluid Solvation Theory)— WaterMap and GIST

Lazaridis (1998) [4] developed IFST (inhomogeneous fluid solvation theory). Young et al., applied Lazaridis' theory to develop the Watermap solvation [5, 6] site-based approach application of IFST toward identifying displaced water sites that enhanced ligand binding. Nguyen [7] developed GIST (Grid Inhomogeneous Solvation Theory), a grid-based application of IFST. Watermap and GIST are examples of explicit solvation simulation models that score the stability of specific water sites within the active sites of proteins. These methods are slower compared to approximate methods that use solvent continuum models, which include only the high dielectric polarization term for water.

Watermap, (implemented within the Glide XP docking commercial software package from Schrodinger, <http://www.schrodinger.com/WaterMap.php>), calculates the entropy of bound water due to the local environment. It is a post-MD (molecular dynamics) trajectory analysis method that uses the OPLS (Optimized Potentials for Liquid Simulations, WL Jorgensen) all atom force field and the TIP4P water model. The local thermodynamic properties of active-site solvent can be computed with explicitly solvated MD simulations to sample the active-site solvent distribution. During MD water molecules are kept track of and the location and orientation of each water molecule is retained in a density profile. Water positions are clustered based on the density profile obtained from the MD simulation. The interaction energy between each water and system is calculated and the entropy of each water molecule estimated. Then active-site solvent distributions can be clustered into high occupancy 1Å spheres, which denote hydration sites of the active site cavity. The average system interaction energy and excess entropy term for water in each hydration site is then calculated and the system interaction energy of the hydration site can be compared with the bulk solvation to estimate the enthalpy cost of water transfer from the hydration active site to the bulk.

Watermap has effectively been used to identify druggable binding sites as these sites usually have a large number of unstable waters [8]. For example WaterMap studies of factor Xa ligand binding. Watermap hydration sites correlated with the experimental SAR studies for fXa ligands. Solvent analysis indentified three enthalpic unfavorable hydration sites, which agreed with experimental data that this sites bound hydrophobic groups [6]. Watermap was tested with a set of 28 ligands extracted from solved crystal structures of factor Xa and the delta G calculated correlated well with delta G experimental.

## **1.2 Implicit Solvation (Solvent Continuum Models)-MM-PB/SA and GB-PB/ SA**

Continuum (implicit) Models are low-level theoretical approaches, which are more approximate and empirical but incur less computational cost than explicit solvent models. Implicit solvent models uncouple polar and nonpolar interactions, assume linear and local solvent polarization, and therefore do not represent a true physical interpretation. In implicit solvent models water is treated as a continuum electrostatic equation to describe polar solvation (continuous homogenous polar liquid) with dielectrics largely based on the Poisson-Boltzman equation. The Poisson-Boltzman Equation is often used to calculate ligand solvation and estimate ligand binding free energies by calculating the total energy of the protein-ligand complex and then subtracting the solvated energy of the protein and ligand separately to give the binding free energy estimates. This commonly used protocol is referred to as the MM-PB/SA method—Molecular Mechanic-Poisson-Boltzmann/Surface Area. Popularized by Kuhn and Kolman, [9] the high computational cost of PB (Poisson Boltzman) caused the development of the GB—generalized Born method. MM-GB/SA has been shown capable of capturing the experimental binding energies [10]. The key weakness of implicit solvation models are their poor description of water-mediated interactions, especially the directionality of water hydrogen bonds, additionally, they treat the nonpolar contribution to solvation approximately.

The choice of an internal dielectric coefficient is important. Biomolecular dielectric coefficients should be 2–4 but low dielectric models are not good for MM/PBSA calculations where molecular flexibility is explicitly modeled through conformational sampling. Dielectric coefficients 4–20 are used for biomolecular applications with lower values being more successful for modeling protein-ligand interactions. The dielectric interface must also be defined, usually as the molecular Connolly surface used to parameterize the biomolecular solvation calculations. Often molecular, Van der Waals surfaces are used sometimes with Gaussian smoothing. The Poisson equation-includes only the influence of the solvent on the electrostatic properties of solute; however, mobile ions can also play a part in electrostatics and solvation. Finite difference methods are the most frequently used numerical methods to solve the Poisson-Boltzman equation for biomolecular electrostatics. The following are all methods which are finite difference-based commonly used Poisson-Boltzmann solvers—APBS (Adaptive Poisson-Boltzmann Solver), DelPhi, MEAD, UHBD, ZAP, PBEQ in CHARMM, PB Solver in AMBER, MIB. SZMAP (commercial software from OpenEye <http://www.eyesopen.com/SZMAP>[11]) is an example of a method based on a semi continuum solvation model. PBSA and GBSA methods are implemented and available within the AmberTools (<http://ambermd.org/doc12/Amber14.pdf>).

### 1.3 Intermediate Approximate Methods—Solvent Potential Methods-3D- RISM

Another alternative theoretical method uses IETS (integral equation theories) to simplify the all atom description of explicit solvation into a probabilistic treatment of solvent and solute distributions. These methods use less computational resources than explicit solvent but are a more detailed description than continuum models. IETS can predict the 3D spatial organization of solvent density around large molecules and thermodynamic solvation quantities. 3D-RISM[12, 13] (three-dimensional reference interaction site model) calculates an approximate average solvent distribution around rigid solute and computes the hydration free energy (HFE) faster than molecular simulations methods. 3D-RISM techniques can compute the thermodynamic effects of water reorganization within a binding pocket without the need to use explicit simulations and can help localize the regions of organized solvent and free energy and have favorable applications for ligand optimization.

3D-RISM does have a problem with the thermodynamics of hydrophobics and does not compute HFE of organics well. The 3D-RISM method is relatively accurate at predicting electrostatic HFE without correction but requires modification of the nonpolar contribution, which originates in the solute water cavity but doesn't include size effect of water molecules in an active site. For 3D RISM one only needs solute and solvent potential parameters as input from molecular two-body additive force fields, such as AMBER, or CHARMM and bulk solvent density, temperature and composition. 3D RISM equations derive the approximate density distribution functions and (DCF) direct correlation functions for each hydrogen and oxygen in water (the three dimensional analog of the radial distribution function). 3D RISM keeps the orientation dependence of solute molecules, which is necessary to describe solvation of large molecular solutes. 1D-RISM is less computationally costly than the 3D method; however, the 3D-RISM method also provides correct dielectric properties of polar solvents. The Chemical Computing MOE 3D RISM is a commercially available implementation of this method [http://www.chemcomp.com/MOE-Structure\\_Based\\_Design.htm](http://www.chemcomp.com/MOE-Structure_Based_Design.htm) [12].

A recent method developed by the research group at GlaxoSmithKline (GSK) combines elements of the IFST and 3D-RISM. The Group at GSK calls their method SPAM [14] which is a statistical mechanics-based approach which estimates free energy difference between protein-bound and bulk water. The SPAM method uses explicit solvent molecular simulations for discrete hydration sites at the water protein interface and computes local free energy of water and site specific interactions. This method provides a qualitative estimate of the thermodynamics of water in hydration sites that agrees well with SAR (structure activity data) and known hot spots. The IFST-based approach estimates the local thermodynamic solvation properties, and interaction energy contribution due to water binding and entropy penalty due to the

ordering of the bulk. In IFST the entropy is calculated by expanding the sum over solute-solvent and solvent-solvent microscopic rotational and translational states. SPAM obtains the entropy contribution from MD simulations. Instead of obtaining states of water from translational and rotational degrees of freedom to obtain entropy, SPAM calculates the distribution of interaction energies between the water and surrounding solute and solvent molecules at a given time. The GSK group has applied their SPAM computational technique to the correct water placement in HIV protease and Hen Egg White Lysozyme. SPAM has provided qualitative estimates that have correlated with experimental SAR observations, for example for HIV1 protease favorable ligand binding hot spots in the protein can be associated with water displaced from the receptor-ligand (causing favorable enthalpy and entropy changes) interaction site or interface to the bulk.

---

## 2 Materials

### Available Software

1. SPC Water and TIP3p, TIP4p, TIP5P solvation models can be used to solvate proteins in GROMACS, [http://www.gromacs.org/Documentation/How-tos/TIP3P\\_coordinate\\_file](http://www.gromacs.org/Documentation/How-tos/TIP3P_coordinate_file); VMD <http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/> AMBER Tools <https://www.cgl.ucsf.edu/chimera/docs/ContributedSoftware/solvate/solvate.html>
2. Watermap (Commerical Software from Schrodinger) <https://www.schrodinger.com/WaterMap.php>
3. MM=GB/SA and MM=PB/SA(AMBER and AmberTools) <http://ambermd.org/tutorials/advanced/tutorial3/>
4. SZMAP (Commercial Software from OpenEye) <http://www.eyesopen.com/SZMAP>
5. APBS (Adaptive Poisson-Boltzmann Solver) <https://sites.google.com/a/poissonboltzmann.org/software/apbs>
6. DelPhi [http://wiki.c2b2.columbia.edu/honiglab\\_public/index.php/Software:DelPhi](http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:DelPhi)
7. MEAD(Macroscopic Electrostatics with Atomic Detail) <http://www.teokem.lu.se/~ulf/Methods/mead.html>; <http://stjuderesearch.org/site/lab/bashford/>
8. GROMACS [http://www.gromacs.org/Documentation/How-tos/TIP3P\\_coordinate\\_file](http://www.gromacs.org/Documentation/How-tos/TIP3P_coordinate_file)
9. SPAM in AmberTools <http://archive.ambermd.org/201401/0290.html>
10. UHBD <http://gilsonlab.umbi.umd.edu>

11. ZAP (electrostatics solver in Commerical OpenEye software) <http://www.eyesopen.com/zap-tk>
12. PBEQ (in CHARMM) <http://www.charmm-gui.org/?doc=input/pbeqsolver>
13. 3D-RISM (Commerical Software Chemical Computing Group (CCG)) [http://www.chemcomp.com/MOE-Structure-Based\\_Design.htm#SolventAnalysiswith3D-RISM](http://www.chemcomp.com/MOE-Structure-Based_Design.htm#SolventAnalysiswith3D-RISM) and [http://dansindhikara.com/Tutorials/Entries/2012/1/1\\_Using\\_3D-RISM\\_and\\_PLACEVENT.html](http://dansindhikara.com/Tutorials/Entries/2012/1/1_Using_3D-RISM_and_PLACEVENT.html)

---

### 3 Methods

#### 3.1 3D-RISM

Dan Sindhikara [13] has made available a tutorial and his code implementing the 3D-RISM water solvation algorithm within AmberTools14 (<http://ambermd.org/doc12/Amber14.pdf>.) He outlines all the methods and steps on his website using code freely available. [http://dansindhikara.com/Tutorials/Entries/2012/1/1\\_Using\\_3D-RISM\\_and\\_PLACEVENT.html](http://dansindhikara.com/Tutorials/Entries/2012/1/1_Using_3D-RISM_and_PLACEVENT.html), <https://sites.google.com/site/dansindhikara/Home/software/placement/tutorial>, [http://dansindhikara.com/Software/Entries/2012/6/22\\_Placement\\_New.html](http://dansindhikara.com/Software/Entries/2012/6/22_Placement_New.html)

1. Parameterize a downloaded PDB structure file
2. Prepare solvent—use SPC water with ions for initial solvation
3. Run 3D-RISM
4. Placevent for water distribution analysis
5. SANDER minimization and simulation

Sindhikara demonstrated accurate predicted water placement in HIV protease [15].

#### 3.2 GROMACS

G\_mmpbsa is an implementation of the MM\_PBSA method by the Open Source Drug Discovery Consortium using GROMACS (Molecular Simulations force field) and the APBS package, which compares favorably with the AMBER MM-PBSA method [16]. This method is freely available [http://rashmikumari.github.io/g\\_mmpbsa/](http://rashmikumari.github.io/g_mmpbsa/). The MM-PBSA method calculates the binding energy and also available with this software are scripts to obtain the binding energy and energetic contribution of each residue. The binding energy consists of three energetic terms, (a) potential energy in vacuum, (b) polar-solvation energy, and (c) nonpolar-solvation energy.

#### 3.3 SPAM

(A Simple Approach for Profiling Bound Water Molecules)[14], a simple way to profile local solvation is now part of Ambertools (AmberTools version 12 and later).

The SPAM code has two components, a Python driver (SPAM.py) and modifications to the cpptraj AmbersTools12 (and later versions)

program module. The SPAM energy evaluations are done with the NAMD software, (because this functionality is not yet part of the Amber software). Prerequisites to running SPAM are therefore AmberTools12 and NAMD installed on your computer.

The SPAM implementation computes the location of hydration sites around the target protein from a multi-nanosecond explicit solvent MD simulation. All steps prior to the water interaction energy calculation are in the cpptraj module AmberTools 12 suite.

SPAM can find the interaction energy of water molecules that travel through the hydration sites of interest. The protein conformation is restrained while solvent molecules can move freely so the system will converge easily and more rapidly where hydration sites are located by identifying the peaks in the computer water density map.

SPAM procedure used by the GSK group:

1. System preparation: water and ligands removed from protein PDB structure.
2. Maestro Protein preparation used protonation states assigned, and the hydrogen bond network optimized.
3. Parm99SB and gaff AMBER force fields used AM1-BCC charges calculated for docking inhibitor.
4. MD simulations with NAMD 2.8 using PME grid spacing 1 Å, 10 Å cutoff with 8 Å switching function –NVT 5,000 step min and 30 ps constant temperature (T=300K) and 1 atm pressure equilibration, followed by 200 ps NPT molecular dynamics followed by production run sampling with a hydration shell for 10 ns with 2fsec time steps. This simulation was able to identify water-binding sites.
5. SPAM analysis—water density map calculated as a time average, with a 5 Å grid using VMD 1.8.7 VolMap—each hydration site of interaction of water molecules calculated using pair Interaction in NAMD.
6. SPAM free energy for hydration site calculated.

### **3.4 Ligand-Docking (Scoring) Improvements**

Docking programs use scoring functions; however, it is a challenge to produce accurate scoring functions, which estimate ligand-binding affinities. The PPC method (polarized protein –specific charge model) included in MM/PBSA was used to rescore ligand binding poses to include the bridging water molecules that were found to play a significant role in determining the protein–ligand binding modes [17]. The accuracy of the MM/PBSA method for docking and predicting ligand-binding affinity relies on the force field accuracy; however, nonpolarizable force fields, such as CHARMM and AMBER, do not accurately represent the protein electrostatics environment. PPC described the polarized electrostatic state of the protein and therefore gives a more accurate description of the electrostatic interactions between the protein and ligand and agrees better with experimental data. Only a few scoring functions take

water into consideration in protein–ligand docking and include binding water molecules in molecular docking. The bridging water molecules were treated as part of the receptor in this method. PPC polarized protein specific charges calculated the electron structure of biomolecules and the continuum dielectric model for solvent. The PPC charge model and MM/PBSA calculations were applied with the Amber ff 99SB on ranking protein–ligand docked poses. The bridging waters need to be explicitly included in the calculation (treated as part of the receptor).

1. Prepare Protein–Ligand complexes from PDB structure coordinates-hydrogens added using Leap AMBER 11 module, all amine groups fully protonated.
2. Partial charges of protein assigned with Amber ff99SB force field.
3. Hydrogens added to ligand and geometry of ligand optimized at HF/6-31G\* level and ANTECHAMBER used to obtain ligand force field parameters. All structures had bridging water molecules.
4. Autodock used to create ensemble of docked conformations for each ligand (bound to receptor).
5. Then 1,500 steps steepest descent followed by 1,500 steps conjugate gradient minimization using AMBER11 Sander.
6. Next step, PPC charges are fitted to electrostatic potentials by QM calculations. Minimized structure used to calculate PPC charges by the MFCC-PB computational protocol –Poisson-Boltzman Delphi.
7. Solvent to calculate induced charges on solute-solvent interface with probe radius 1.4 Å. Solvent dielectric constant set to 80 and grid density of 4 grids/ angstrom used to numerically solve Poisson-Boltzman equation.
8. QM calculation of protein fragments performed with DFT theory at B3LYP/6-31G\*. Molecular Dynamics simulations with TiP3P water and counterions and Amber ff99SB force field.
9. MM/PBSA calculation of the binding affinity.

With molecular docking with structural waters in the binding pocket, Autodock still could not predict the native structure poses. Rescoring the binding free energies using the MM/PBSA and structural waters, including the bridging waters, the MM/PBSA with Amber ff99SB method recognizes the correct binding for some structures. The MM/PBSA and PPC model recognizes all the native binding poses since the polarized state of the protein provides accurate electrostatic interactions and near native structure.

### **3.5 Applications to FBDD-**

Fragments are smaller and less complex than lead compounds and therefore have fewer degrees of freedom for virtual screening and docking. Additionally, fragments have greater solubility and polarity

than regular molecular hits and lower binding affinities. Docking scoring functions sometimes neglect solvation, including solvation that would make fragment docking scoring more accurate since solvation is a greater proportion of the binding free energy in fragments than in larger lead molecules. Docking fragments rescored with MM-PBSA, which gave significantly better docking results [18].

### 3.6 Conclusion

1. Active site waters play a significant role and need to be considered in the development of computational docking and ligand binding site studies.
2. Computational and theoretical methods to describe the energetics of water and solvent interactions at ligand binding site are improving in their ability to predict and correlate with experimental binding data.

## References

1. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1993) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 99: 926–935
2. Toukan K, Rahman A (1985) Molecular-dynamics study of atomic motions in water. *Phys Rev B* 31:2643–2648
3. Kästner J, Senn HM, Thiel S, Otte N, Thiel W (2006) QM/MM free-energy perturbation compared to thermodynamic integration and umbrella sampling: application to an enzymatic reaction. *JCTC* 2 2:452–461
4. Lazaridis T (1998) Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. *J Phys Chem B* 102:3531–3541
5. Young T, Abel R, Kim B, Berne BJ, Friesner RA (2007) Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc Natl Acad Sci U S A* 104:808–813
6. Abel R, Young T, Farid R, Berne BJ, Friesner RA (2008) Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J Am Chem Soc* 130:2817–2831
7. Nguyen CN, Young TK, Gilson MK (2012) Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit [7] uril. *J Chem Phys* 137:044101–044117
8. Beuming T, Che Y, Abel R, Kim B, Shanmugasundaram V, Sherman W (2012) Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins* 80:871–883
9. Kuhn B, Kollman PA (2000) A ligand that is predicted to bind better to avidin than biotin: insights from computational fluorine scanning. *J Am Chem Soc* 122:3909–3916
10. Guimaraes CRW, Cardozo M (2008) MM-GB/SA resourcing of docking poses in structure-based lead optimization. *J Chem Inf Model* 48:958–970
11. SZMAP 1.2.0.7: OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>. (2013)
12. Truchon JF, Pettitt BM, Labute P (2014) A cavity corrected 3D-RISM functional for accurate solvation free energies. *J Chem Theory Comput* 10:934–941
13. Sindlikara DJ, Hirata F (2013) Analysis of biomolecular solvation sites by 3D-RISM theory. *J Phys Chem B* 117:6718–6723
14. Cui G, Swails JM, Manas ES (2013) SPAM: a simple approach for profiling bound water molecules. *J Chem Theory Comput* 9:5539–5549
15. Sindlikara DJ, Yoshida N, Hirata F (2012) An algorithm for prediction of explicit solvent atom distribution—application to HIV-1 protease and F-ATP synthase. *J Comput Chem* 33:1536–1543
16. Kumari R, Kumar R, Open Source Drug Discovery Consortium, Lynn A (2014) g\_mmmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model* 54:1951–1962
17. Liu J, He X, Zhang JZH (2013) Improving the scoring of protein-ligand binding affinity by including the effects of structural water and electronic polarization. *J Chem Inf Model* 53:1306–1314
18. Kawatkar S, Moustakas D, Miller M, Joseph-McCarthy D (2012) Virtual fragment screening: exploration of MM-PBSA re-scoring. *J Comput Aided Mol Des* 26:921–934

# Chapter 2

## Binding Site Druggability Assessment in Fragment-Based Drug Design

Yu Zhou and Niu Huang

### Abstract

Target druggability refers to the propensity that a particular target is amenable to bind high-affinity drug-like molecules. A robust yet accurate computational assessment of target druggability would greatly benefit the fields of chemical genomics and drug discovery. Here, we illustrate a structure-based computational protocol to quantitatively assess the target binding-site druggability via in silico screening a fragment-like compound library. In particular, we provide guidelines, suggestions, and critical thoughts on different aspects of this computational protocol, including: construction of fragment library, preparation of target structure, in silico fragment screening, and analysis of druggability.

**Key words** Druggability assessment, Fragment screening, Molecular docking, MM-GB/SA rescoring, Hit rate

---

### 1 Introduction

Successful drug development requires a disease target of both biological relevance and chemical tractability. With the completion of the human genome, we now have unprecedented access to large numbers of potential therapeutic targets. The question that arises is which specific protein targets can be modulated by a drug-like molecule. Druggability (i.e., propensity that a particular target is amenable to bind high-affinity drug-like molecules) assessment in the process of target selection would reduce drug discovery attrition and put effort on those targets most likely to lead to therapeutic intervention [1].

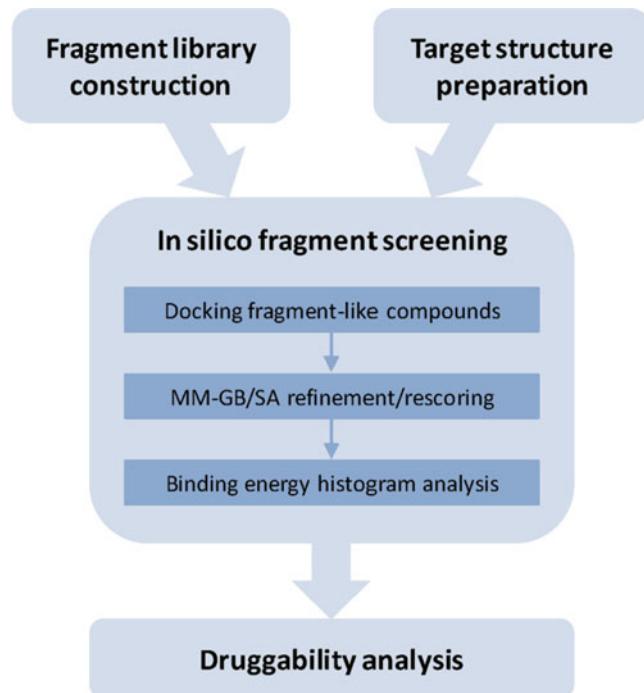
The first step in evaluating the druggability of a target is to identify the presence of binding pockets with suitable size, shape, and composition to accommodate drug-like molecules. Many approaches for this purpose have been developed that are generally classified as geometry-based [2–5], information-based [6, 7], and energy-based algorithms [8, 9]. Benchmarking studies using

training set data extracted from the Protein Data Bank (PDB), most approaches have demonstrated to correctly detect the true ligand-binding sites. However, the presence of a suitable protein pocket is necessary but not sufficient to guarantee potent binding of drug-like small molecules.

The more difficult step is to quantitatively predict the drug-gability index of a given binding site. Early studies have predicted target druggability on the basis of sequence and structure homology to known drug targets [10, 11]. However, not all members of the same protein family are equally druggable [12]. More importantly, such methods cannot be used to assess drug-gability of novel target families. Recently several structure-based target druggability methods have been developed and validated against a set of reference targets where the degree of tractability is known. These methods provide quantitative assessments of druggability using physicochemical descriptors derived from the ligand binding pockets and apply techniques as varied as bio-physical modeling [13], linear regression [14, 15], and support vector machines [16].

Hajduk et al. made a seminal contribution by demonstrating that experimental hit rates from the heteronuclear-NMR-based fragment screening could serve as an effective druggability index within a set of 23 protein targets containing 28 different binding sites [17]. Furthermore, they derived a linear regression model to fit the experimentally measured hit rates to physicochemical descriptors of these 28 binding pockets. Applying an appropriate cutoff, this model was assessed using an additionally assembled binding-site dataset, and 33 out of 35 known drug-like ligand-binding sites were correctly identified. Being essentially analog to the NMR-based fragment screening, an *in silico* fragment screening protocol was also developed to assess target binding-site druggability [18]. It makes use of a molecular mechanics-based scoring method for the protein–ligand interaction and the obtained virtual hit rates were demonstrated to correlate with the hits rate measured experimentally from the NMR-based screening method. This protocol can be employed to distinguish known druggable and non-druggable targets, and it is generally applicable without relying on any assembled training data set that potentially extends its capacity toward unexplored target space.

In this chapter, we illustrate the computational details of this *in silico* fragment screening protocol for target druggability assessment (*see* Fig. 1 for a schematic overview). We outline the criteria for the construction of fragment library, discuss the method for the preparation of target structure, and describe the procedure for carrying out the *in silico* fragment screening. Finally, we discuss the druggability analysis from the virtual screening results.



**Fig. 1** Schematic illustration of druggability prediction via fragment-based docking and scoring approach

---

## 2 Materials

The druggability assessment protocol entails building a fragment-like compound library and performing in silico fragment screening experiments, which could be carried out by means of a variety of Web servers and software. The programs listed here are merely the ones used as examples for illustrating this procedure. The diverse set of fragments is selected from the fragment-like subset of the ZINC database [19, 20]. The DOCK 3.5.54 program [21, 22] is used to dock the fragment database into the protein binding site. The Protein Local Optimization Program (PLOP) [23–25] is used to perform MM-GB/SA refinement and rescoring.

---

## 3 Methods

### 3.1 Fragment Library Construction

1. Extract compounds from the fragment-like subset of the ZINC database (*see Note 1*).
2. Eliminate fragments with more than 15 heavy atoms (*see Note 2*).

3. Calculate feature key fingerprints using CACTVS [26], and perform the fingerprint-based similarity analysis with a modified version of the program SUBSET [27] to reduce redundancy of the fragment library (*see Note 3*).

### 3.2 Target Structure Preparation

1. Select one or more representative structures for the protein target (*see Note 4*).
2. Determine the ligand binding pocket (*see Note 5*). Identify cofactors, metal ions, and structural waters in the target protein and treat them as part of the protein if they are involved in ligand binding.
3. Add hydrogen atoms to the protein. Assign proper protonation states for binding-site residues and optimize the orientations for polar hydrogen atoms using PLOP (*see Note 6*).

### 3.3 In Silico Fragment Screening

The in silico screening protocol employs a physics-based hierarchical scoring method which consists of two steps: predicting the binding poses of ligands using a docking program, and then refining and rescoring those protein–ligand complexes using a more computationally intensive molecular-mechanics based energy function [28, 29]. This protocol uses a high-throughput docking program to initially orient and score the ZINC fragment-like compounds in the binding site, and subjects the best single docking pose for each docked compound to a rescoring stage in which the ligand is fully minimized inside the binding site and the binding energy is estimated with an all-atom molecular mechanics force field combined with an implicit solvent model. Finally the results of all compounds are analyzed based on the binding energy distribution.

#### 3.3.1 Docking Fragment-Like Compounds Library

1. Identify binding site residues within a certain range (e.g., 12 Å) away from any heavy atom of the crystallographic ligand or the residues used to define the site, using the program FILT (part of the UCSF DOCK suite).
2. Calculate the solvent-accessible molecular surface [30] of the protein binding site with the program DMS [31] using a probe radius of 1.4 Å.
3. Generate receptor-derived spheres with the program SPHGEN (part of the UCSF DOCK suite) [32], in combination with the ligand-derived spheres if necessary (*see Note 7*).
4. Set the grid box dimensions with edges 15 Å beyond the matching spheres initially. Then refine the box dimensions to maximize the coverage of the protein without exceeding 2 million grid points at a predefined grid resolution (three points per angstrom by default). Finally, four scoring grids are generated: an excluded volume grid using DISTMAP [33], a united

atom AMBER-based van der Waals potential grid using CHEMGRID [33], an electrostatic potential grid using DelPhi [34], and a solvent occlusion map using the program SOLVMAP [35].

5. Perform docking with DOCK 3.5.54, a flexible-ligand method that uses a force-field-based scoring function. Ligand conformations are scored on the basis of the total docking energy ( $E_{\text{tot}} = E_{\text{ele}} + E_{\text{vdw}} - \Delta G_{\text{lig-solv}}$ ), which is the sum of electrostatic ( $E_{\text{ele}}$ ) and van der Waals interaction energies ( $E_{\text{vdw}}$ ), corrected by the partial ligand desolvation energy ( $\Delta G_{\text{lig-solv}}$ ).
  6. Save a single docking pose with the best total energy score for each docked molecule for the next stage of scoring (*see Note 8*).
1. Generate OPLS force field parameter for each molecular compound and cofactor (if present), using IMPACT (part of the Schrödinger suite).
  2. Submit the free ligand, free protein and docked protein–ligand complex to multi-scale Truncated Newton (MSTN) energy minimization [25] in all-atom OPLS force field [36, 37] and Generalized Born (GB) solvent [38, 39] using PLOP (*see Note 9*).
  3. Calculate the binding energy ( $E_{\text{bind}} = E^{\text{RL}} - E^{\text{L}} - E^{\text{R}}$ ) by subtracting the energies of the optimized free ligand in solution ( $E^{\text{L}}$ ) and the free protein in solution ( $E^{\text{R}}$ ) from the optimized protein–ligand complex’s energy in solution ( $E^{\text{RL}}$ ) (*see Note 10*).

### 3.3.2 MM-GB/SA Refinement and Rescoring

1. Report the energy scores distribution for the protein target.

### 3.3.3 Histogram Analysis of Energy Score

## 3.4 Druggability Analysis

1. Compute the “hit rate” for the in silico screening based on a chosen energy cutoff value (-40 kcal/mol) (*see Note 11*).
2. Calculate the druggability score which is defined as  $\log(\text{hit rate})$ .
3. Compare the druggability score with the cutoff value of 0.36 to classify the assessed target as druggable or non-druggable (*see Note 12*).

---

## 4 Notes

1. Fragments are molecules of low complexity, which sample chemical space exponentially more effectively than drug-sized molecules. Different estimates exist of the size of chemical space. Here, the fragment-like subset of the ZINC database (version 6, December 2005) contains 49,134 compounds with relatively low molecular weight ( $\text{MW} \leq 250$ ), few rotatable

bonds ( $\text{RB} < 3$ ), low hydrophobicity ( $-2 < \log P < 3$ ), and weak hydrogen bonding potentials ( $\text{HB}_{\text{donor}} < 3$  and  $\text{HB}_{\text{acceptor}} < 6$ ).

2. Kuntz et al. observed that the maximal binding free energy increases more slowly for ligands containing more than 15 heavy atoms [40]. Therefore, fragments with more than 15 heavy atoms were eliminated. This filter reduced the library size to 32,717 molecules.
3. Representative structures were selected for each structural cluster with Tanimoto coefficient ( $T_c$ ) less than 0.9 to other clusters. This further reduced the library to 11,129 diverse molecules. To assess any potential bias resulting from the diversity-based filtering, redo the screening using 32,717 ZINC fragment-like compounds for the training dataset, leads to very similar energy distributions.
4. Targets may have multiple crystal complex structures available and some display significant side-chain movement upon binding to different ligands [41]. In most cases, we found that the changes of the histograms of energy scores and the druggability scores calculated from them are remarkably small when using different crystal structures. Nevertheless, multiple conformations are recommended for the binding sites with large structural variation, especially for the protein–protein interaction (PPI) interfaces. Applying our protocol, specific druggable conformations could also be identified.
5. The identification of the protein binding pocket is straightforward for ligand-bound complex structures. However, the binding site is not known from a 3-D structure or from other experimental data, a “suitable” pocket is required to be detected firstly by pocket detection programs or virtual inspection.
6. Ideally, the target protein should be prepared as if the crystal ligand was absent, as adjusting the protein to favor crystal ligands is a source of bias.
7. Spheres are generated to fill the binding site. Matching spheres required for the orientation of the ligand within the binding site are obtained by augmenting the ligand-derived spheres with receptor-derived spheres. By default, spheres furthest away from ligand-derived spheres, furthest from the centroid of the remaining spheres, too close to receptor atoms, or too close to each other are removed iteratively until the total number of sphere is 35 or less. However, for large binding surfaces like protein–protein interfaces, we use a maximum of 120 matching spheres to ensure adequate ligand sampling.
8. One major limitation of the current protocol is that it relies entirely on the docking algorithm to identify the correct binding pose. A simple extension of this protocol is to subject

a small number of dissimilar binding poses to minimization in the MM-GB/SA rescoring step and use the most favorable binding energy for rank-ordering ligands. Therefore, multiple (usually hundreds of) docking poses could be saved in docking stage and subjected for structural descriptor-based filtering and KGS-penalty function-based conformational clustering [42]. Tens of poses might be finally obtained for next MM-GB/SA rescoring.

9. The molecular mechanics forces are divided into short-range (bond, angle, torsion, and local non-bonded) and long-range components, with the long-range forces updated only intermittently. The algorithm is also optimized for minimizations with GB solvent that increases the computational expense by only a factor of ~3 relative to the vacuum. Thus, this scoring approach accounts for accurate and efficient calculations of ligand–protein interaction energies, the ligand/receptor desolvation, and to a lesser extent, ligand strain energies. In this work, the protein was kept rigid during protein–ligand minimization to reduce the computational expense.
10. Accurate free energy calculations depend on a proper balance of many different energetic components. The MM-GB/SA rescoring method strikes a balance between computational speed and accuracy, and in particular neglects entropic loss and protein flexibility. Empirically scaling certain energy components as a post-rescoring process, in a manner similar to LIE scheme, may be useful to compensate for some of these limitations [43]. It has been suggested that the MM-GB/SA scoring function underestimates the nonpolar binding contributions to the free energy of binding [28]. In this study, we empirically scaled the van der Waals energy component by a factor of 2.
11. This cutoff value was empirically chosen to maximally differentiate druggable and non-druggable binding site. We visually inspected the energy distributions for the 13 druggable binding sites and 11 non-druggable binding sites in Hajduk et al. training data set and explored the effect of varying the cutoff with respect to differentiating between druggable and non-druggable binding sites. We found the correlation between the docking screening hit rates and the NMR screening results is relatively insensitive to the value of the energy cutoff within a certain range (from -40 to -34 kcal/mol). In this work, an energy cutoff of -40 kcal/mol was used for computing the in silico hit rate.
12. The calculated druggability scores correlate reasonably well with the NMR-based fragment screening results. Hajduk et al. defined binding sites as “highly druggable” if they have a experimental  $\log(\text{hit rate}) > -1.0$ . The corresponding value of computational  $\log(\text{hit rate})$  is 0.36, and we used this value to

classify proteins as druggable or non-druggable in this work. Although Hajduk et al. distinguish between “highly druggable” and “moderately druggable,” we use a simple binary classification for simplicity. Nevertheless, the higher druggability score a target is assigned, the more druggable it might be.

## Acknowledgement

The Chinese Ministry of Science and Technology “973” Grant 2011CB812402 (to N.H.) is acknowledged for financial support, Shoichet Lab at UCSF for the DOCK3.5.54 program and Jacobson Lab at UCSF for PLOP.

## References

- Fauman EB, Rai BK, Huang ES (2011) Structure-based druggability assessment—identifying suitable targets for small molecule therapeutics. *Curr Opin Chem Biol* 15:463–468
- Brady GP Jr, Stouten PF (2000) Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* 14:383–401
- Hendlich M, Rippmann F, Barnickel G (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J Mol Graph Model* 15(359–363):389
- Laskowski RA (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* 13(323–330):307–328
- Liang J, Edelsbrunner H, Woodward C (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci* 7:1884–1897
- Soga S, Shirai H, Kobori M, Hirayama N (2007) Use of amino acid composition to predict ligand-binding sites. *J Chem Inf Model* 47:400–406
- Stuart AC, Ilyin VA, Sali A (2002) LigBase: a database of families of aligned ligand binding sites in known protein sequences and structures. *Bioinformatics* 18:200–201
- An J, Totrov M, Abagyan R (2004) Comprehensive identification of “druggable” protein ligand binding sites. *Genome Inform* 15:31–41
- Glaser F, Morris RJ, Najmanovich RJ, Laskowski RA, Thornton JM (2006) A method for localizing ligand binding pockets in protein structures. *Proteins* 62:479–488
- Hopkins AL, Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* 1:727–730
- Blundell TL, Sibanda BL, Montalvao RW, Brewerton S, Chelliah V, Worth CL, Harmer NJ, Davies O, Burke D (2006) Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. *Philos Trans R Soc Lond B Biol Sci* 361:413–423
- Fauman EB, Hopkins AL, Groom CR (2003) Structural bioinformatics in drug discovery. *Methods Biochem Anal* 44:477–497
- Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, Salzberg AC, Huang ES (2007) Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* 25:71–75
- Sheridan RP, Maiorov VN, Holloway MK, Cornell WD, Gao YD (2010) Drug-like density: a method of quantifying the “bindability” of a protein target based on a very large set of pockets and drug-like ligands from the Protein Data Bank. *J Chem Inf Model* 50:2029–2040
- Halgren TA (2009) Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* 49:377–389
- Nayal M, Honig B (2006) On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins* 63:892–906
- Hajduk PJ, Huth JR, Fesik SW (2005) Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* 48:2518–2525
- Huang N, Jacobson MP (2010) Binding-site assessment by virtual fragment screening. *PLoS One* 5:e10109

19. Irwin JJ, Shoichet BK (2005) ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45:177–182
20. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012) ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model* 52:1757–1768
21. Lorber DM, Shoichet BK (2005) Hierarchical docking of databases of multiple ligand conformations. *Curr Top Med Chem* 5:739–749
22. Wei BQ, Baase WA, Weaver LH, Matthews BW, Shoichet BK (2002) A model binding site for testing scoring functions in molecular docking. *J Mol Biol* 322:339–355
23. Jacobson MP, Kaminski GA, Friesner RA, Rapp CS (2002) Force field validation using protein side chain prediction. *J Phys Chem B* 106:11673–11680
24. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA (2004) A hierarchical approach to all-atom protein loop prediction. *Proteins* 55:351–367
25. Zhu K, Shirts MR, Friesner RA, Jacobson MP (2007) Multiscale optimization of a truncated newton minimization algorithm and application to proteins and protein-ligand complexes. *J Chem Theory Comput* 3:640–648
26. Ihlenfeldt WD, Takahashi Y, Abe S, Sasaki S (1994) Computation and management of chemical properties in CACTVS: an extensible networked approach toward modularity and flexibility. *J Chem Inf Comput Sci* 34:109–116
27. Voigt JH, Bienfait B, Wang S, Nicklaus MC (2001) Comparison of the NCI open database with seven large chemical structural databases. *J Chem Inf Comput Sci* 41:702–712
28. Huang N, Kalyanaraman C, Bernacki K, Jacobson MP (2006) Molecular mechanics methods for predicting protein-ligand binding. *Phys Chem Chem Phys* 8(44):5166–5177
29. Huang N, Kalyanaraman C, Irwin JJ, Jacobson MP (2006) Physics-based scoring of protein-ligand complexes: enrichment of known inhibitors in large-scale virtual screening. *J Chem Inf Model* 46:243–253
30. Connolly ML (1983) Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221:709–713
31. Ferrini TE, Huang CC, Jarvis LE, Roberts L (1988) The MIDAS display system. *J Mol Graph* 6:13–27
32. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 161:269–288
33. Meng EC, Shoichet BK, Kuntz ID (1992) Automated docking with grid-based energy evaluation. *J Comput Chem* 13:505–524
34. Nicholls A, Honig B (1991) A raid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J Comput Chem* 12:435–445
35. Mysinger MM, Shoichet BK (2010) Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* 50:1561–1573
36. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118:11225–11236
37. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105:6474–6487
38. Gallicchio E, Zhang LY, Levy RM (2002) The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *J Comput Chem* 23:517–529
39. Ghosh A, Rapp CS, Friesner RA (1998) Generalized born model based on a surface integral formulation. *J Phys Chem B* 102:10983–10990
40. Kuntz ID, Chen K, Sharp KA, Kollman PA (1999) The maximal affinity of ligands. *Proc Natl Acad Sci U S A* 96:9997–10002
41. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R (2006) Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem* 49:534–553
42. Peng SM, Zhou Y, Huang N (2013) Improving the accuracy of pose prediction in molecular docking via structural filtering and conformational clustering. *Chin Chem Lett* 24:1001–1004
43. Zhou Z, Madura JD (2004) CoMFA 3D-QSAR analysis of HIV-1 RT non nucleoside inhibitors, TIBO derivatives based on docking conformation and alignment. *J Chem Inf Comput Sci* 44:2167–2178

# Chapter 3

## Generating “Fragment-Based Virtual Library” Using Pocket Similarity Search of Ligand–Receptor Complexes

Raed S. Khashan

### Abstract

As the number of available ligand–receptor complexes is increasing, researchers are becoming more dedicated to mine these complexes to aid in the drug design and development process. We present free software which is developed as a tool for performing similarity search across ligand–receptor complexes for identifying binding pockets which are similar to that of a target receptor. The search is based on 3D-geometric and chemical similarity of the atoms forming the binding pocket. For each match identified, the ligand’s fragment(s) corresponding to that binding pocket are extracted, thus forming a virtual library of fragments (FragVLib) that is useful for structure-based drug design. The program provides a very useful tool to explore available databases.

**Key words** Fragment-based, Drug design, Virtual library, In silico, Pocket similarity, Subgraph mining

---

### 1 Introduction

We present a tool that mine ligand–receptor complexes and generate a library of fragments for a target receptor so it can be used for structure-based drug design, such as Fragment-Based Lead Design (FBLD). FBLD is a computational approach which begins with a small low affinity fragment(s) which bind to the target of interest, followed by a careful construction and optimization of these fragments to end up with a high affinity lead drug. In theory, this is a highly efficient approach for drug design, and it has become enormously popular in the past few years [1–4].

Our method, FragVLib [5], relies on a *Graph*-based representation for interfacial atoms of a ligand–receptor complex. Interfacial atoms are defined as the adjacent receptor and ligand atoms which are within certain cutoff distance. Interfacial atoms are represented by nodes, and distances between them are represented by edges connecting these nodes. Therefore, the resulting interfacial-graph contains a number of nodes representing atoms from the ligand

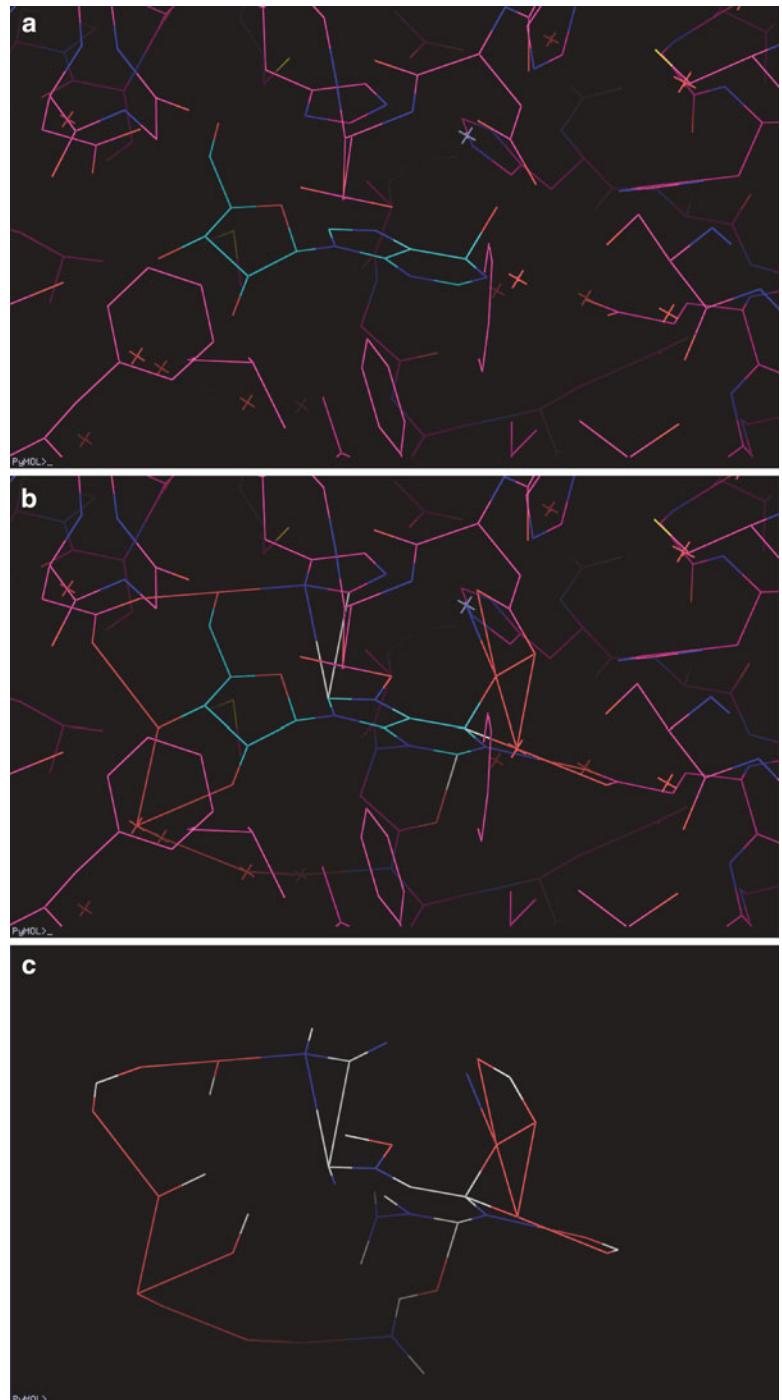
connected by edges to a number of nodes representing atoms from the receptor. Furthermore, the interfacial-graph also includes all the atoms that are covalently bound to the interfacial atoms. These atoms are represented by nodes, and the covalent bonds connecting them to the interfacial atoms are represented by edges (Fig. 1).

We should mention that we make use of the tessellation technique to identify the interfacial atoms. Specifically, we use almost-Delaunay (AD) tessellation [6] which has a unique advantage of incorporating the imprecision of the point coordinates in defining the tessellation patterns. Besides the cutoff distance (*ADdistance*) used to identify adjacent atoms, a threshold value (*ADepsilon*) is used to signify the minimum perturbation needed for an atom to be part of the interfacial atoms. This is important when dealing with bad-resolution ligand–receptor complexes.

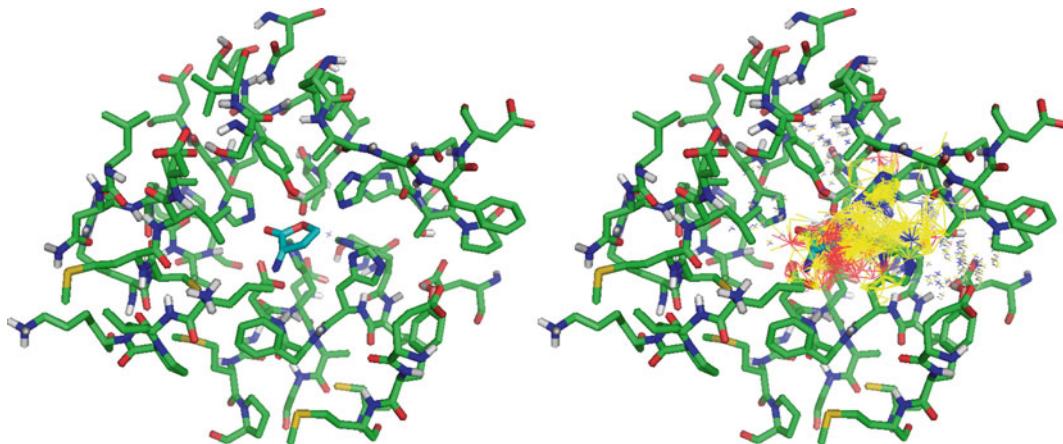
Now let us assume that we have a “target” ligand–receptor complex for which we are interested in designing a lead compound using FragVLib method. Let us also assume that we have a database of X-ray crystallized ligand–receptor complexes, i.e., “native” complexes. First, we will generate the interfacial-graphs for all ligand–receptor complexes involved, i.e., the target complex and all the native complexes.

Now since we have the complexes’ interfaces represented by interfacial-graphs, we can use efficient subgraph match to perform a pocket similarity search between the interfacial-graph of the target complex and the interfacial-graph of each one of the “native” complexes in the database. The match considers all possible subgraphs and is performed over the atoms and bonds composing the receptor side of the interfacial graphs *only*; this is a pocket similarity search, and ligands were only used to define the binding pockets. The match takes into account the chemistry and the 3D geometry of the matching atoms and bonds. The 3D geometry is checked by making sure that the matching atoms superimpose within a user-defined RMSD cutoff value (*dRMSDcutoff*). The user of the tool (FragVLib) can also limit the size of an accepted match (i.e., number of nodes in the matched subgraph) by providing the minimum value (*minMatchSize*) and a maximum value (*maxMatchSize*) for the size.

Every time an accepted subgraph match is identified between the interfacial-graph of the target complex and the interfacial-graph of a native complex, the ligand’s part (atoms and bonds) of the interfacial-graph of the native complex that are only in direct contact with the identified subgraph match is copied into the pocket of the target receptor. When repeating this pocket similarity search using each native complex in the database, we will generate a collection of chemical fragments filling the binding pocket of the target receptor. These fragments constitute the so called “Fragment-based Virtual Library” or FragVLib (Fig. 2).



**Fig. 1** (a) An example of a receptor–ligand complex. (b) The same example after defining interfacial atoms using almost-Delaunay (AD) tessellation. (c) The interfacial atoms and their bonds form the interfacial-graph



**Fig. 2** A picture for the target receptor–ligand complex on the *left side*, and another picture for the receptor after identifying the fragments (virtual library of fragments) using FragVLib, on the *right side*

Finally, for lead design, the user can explore these fragments and perform one of the following: growing them into the depth of the binding pocket; carefully connecting two or more fragments into one compound for optimized potency; or merging two or more fragments in regions of mutual overlap to construct a lead compound [7].

## 2 Materials

The program is written in C++, and it is publicly available freeware; it can be copied and distributed freely. The user manual and the pre-compiled executables can be downloaded by going to the website “<http://www.bioinformatics.org/fraglib>” and installing the file “FragVLib.zip”. It is easy to install (no external libraries) and easy to use as we explain in the next section. After unzipping the file, you will have the following executables (*see Note 1*, and *Note 7*):

- *getIntGraph4Target*
- *getIntGraphs4DB*
- *FragVLib*
- *rmLigHs*
- *rmProHs*
- *rmProWs*
- *getAlmDisGraphMol2*
- *mol2graphXYZ*
- *ADCGAL*
- *ADedgeCGAL*

Notice that all of them run on a Linux operating system. You will have the target receptor–ligand complex for which you would like to design the lead compound in the PDB file format and in MOL2 file format, for the receptor and the ligand, respectively. You will also have a database of native, X-ray crystallized, receptor–ligand complexes in the same file format. You will need a program like *PyMOL* to view the fragments after generating them; you can install it from this website: “<http://www.pymol.org/>”.

---

### 3 Methods

The following are the steps you will need to generate the virtual library of fragments (FragVLib). You need to have all the executable files and your files in one directory.

#### 3.1 Obtaining the Interfacial-Graph for the Target Receptor

The first step in this method is obtaining the interfacial-graph for the target receptor. You should have the target receptor–ligand complex available in MOL2 file format for the ligand, and in PDB file format for the receptor. Then you can type the following command:

```
getIntGraph4Target namesFile ADdistance ADepsilon noW
```

The *namesFile* is a file containing the name (including location) of the ligand’s file, followed by space, followed by name (including location) of the receptor’s file. The *ADdistance* is the maximum distance for two interfacial atoms to be considered in contact, the recommended value is between 3.5 and 5.8 Å. The *ADepsilon* parameter is the maximum perturbation allowed for the location of an atom, the recommended value is between 0.01 and 0.1 Å. Go back to the Subheading 1 for more details about these parameters (*also, see Note 2*). The *noW* is a parameter that, if included, tells the program to ignore water molecules and treat them implicitly (*see Note 3*). If you want water molecules to be part of the interface, simply do not include this parameter. Below are two examples of running the *getIntGraph4Target* program:

```
getIntGraph4Target namesFile 4.25 0.05 noW
```

```
getIntGraph4Target namesFile 4.0 0.01
```

The output of this step will be two files for the atoms and bonds of the target receptor’s interfacial-graph: *Target\_atomsXYZ*, and *Target\_bonds*.

#### 3.2 Obtaining the Interfacial-Graphs for the Database of Complexes

The second step is obtaining the interfacial-graphs for the database of X-ray crystallized (native) receptor–ligand complexes. For each complex, you should have the ligand’s file in MOL2 file format, and the receptor’s file in PDB file format. You need to list the names of all receptor–ligand complexes in one file *namesFile*, such

that each line refers to one complex and contains the name (including location) of the ligand's file, followed by space, and followed by the name (including location) of receptor's file. Then you will type the following command:

```
getIntGraphs4DB namesFile ADdistance ADepsilon noW
```

Make sure you use the same values for parameters *ADdistance* and *ADepsilon* used in previous step when obtaining interfacial-graph for the target receptor. The output of this step will be two files for the atoms and bonds of the interfacial-graphs: *DB\_atomsXYZ\_name*, and *DB\_bonds*.

### 3.3 Generating the Virtual Library of Fragments

Finally, the last step is performing the pocket similarity search between the target receptor's interfacial-graph, and the interfacial-graph for each receptor-ligand complex in the database. A subgraph match will start by running the following command:

```
FragVLib Target_atomsXYZ Target_bonds DB_atomsXYZ_name  
DB_bonds minMatchSize minMatchSize dRMSDcutOff outDir
```

The first four files are the same ones generated in the previous two steps, so you will not have to do anything about them. The *minMatchSize* and *maxMatchSize* is the minimum and maximum size of a matched interface to be accepted (*see Note 4*). The *dRMSDcutoff* is the maximum value for an RMSD of the matching pockets to be accepted as similar pockets, it can take a value from 0.1 to 1.0 Å. Go back to the Subheading 1 for more details about these three parameters (*also, see Note 5*). The *outDir* is the directory where all the generated fragments will be stored in (*see Note 6*). These fragments will constitute the virtual library, and they will be stored in MOL2 file format. You can use *PyMOL* to view the fragments and start the lead design process.

## 4 Notes

1. The program utilizes efficient tools for representing the interfacial atoms of the ligand–receptor complexes, as well as performing the pocket similarity search. However, the major drawback for the method is the fact that it relies on subgraph matching as a way of performing the match searching process. Subgraph mining in the presence of isomorphism is a well-known NP-Complete problem [8] in the field of computer science. Such kind of problems is typically solved using techniques such as: Approximation, Randomization, Parameterization, Restriction, and Heuristic algorithms. Herein, to speed up the searching process, we implemented parameterization, restriction and heuristic algorithms. Parameterization is possible by controlling certain input parameters, such as: *ADdistance*,

*ADepsilon*, *minMatchSize*, *maxMatchSize*, and *dRMSDcutoff*. For example, using short cutoff distances (*ADdistance*) in identifying interfacial atoms will result in interfacial-graphs that are smaller in size, and therefore, faster search is obtained.

2. Short cutoff distances (*ADdistance*) can be used when the target receptor's binding pocket is expected to have interactions such as: hydrogen bond and ion exchange, which occur over short distances. If we expect hydrophobic interactions, which can occur over large distances, higher cutoff values can be used.
3. Water molecules can be included as part of the interface, or they can be omitted and treated implicitly by adding the *noW* parameter. Omitting water molecules will speed up the search process.
4. You can modify the size of the matching binding pockets to search for a smaller binding region in the target receptor by modifying values of *minMatchSize*, and *maxMatchSize*.
5. The RMSD cutoff value (*dRMSDcutoff*) for accepting the matched (superimposed) interfacial-graphs can be used to decide how (geometrically) similar the binding matching binding pockets are.
6. If you decide to run another round of *FragVLib*, make sure you choose a different name for the *outDir*, or delete the one you have.
7. Always make sure you have all the executables (listed in Subheading 2) in the same directory where you are running the program.

## References

1. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* 6:211–219
2. Loving K, Alberts I, Sherman W (2010) Computational approaches for fragment-based and de novo design. *Curr Top Med Chem* 10:14–32
3. Rees DC, Congreve M, Murray CW, Carr R (2004) Fragment-based lead discovery. *Nat Rev Drug Discov* 3:660–672
4. Schneider G, Fechner U (2005) Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov* 4:649–663
5. Khashan R (2012) *FragVLib* a free database mining software for generating “Fragment-based Virtual Library” using pocket similarity search of ligand-receptor complexes. *J Cheminform* 4:18
6. Bandyopadhyay D, Snoeyink J (2007) Almost-Delaunay simplices: robust neighbor relations for imprecise 3D points using CGAL. *Comput Geom* 38:4–15
7. Detering C, Gastreich M, Lemmen C (2010) Leading fragments to lead structures: fragment evolution, merging and core replacement, and docking. *Chem Inform Bull* 62:62
8. Huan J, Wang W, Prins J (2003) Efficient mining of frequent subgraph in the presence of isomorphism. In: Proceedings of the 3rd IEEE international conference on data mining (ICDM), Melbourne, Florida, 19–22 Nov 2003

# Chapter 4

## Virtual Fragment Preparation for Computational Fragment-Based Drug Design

Jennifer L. Ludington

### Abstract

Fragment-based drug design (FBDD) has become an important component of the drug discovery process. The use of fragments can accelerate both the search for a hit molecule and the development of that hit into a lead molecule for clinical testing. In addition to experimental methodologies for FBDD such as NMR and X-ray Crystallography screens, computational techniques are playing an increasingly important role. The success of the computational simulations is due in large part to how the database of virtual fragments is prepared. In order to prepare the fragments appropriately it is necessary to understand how FBDD differs from other approaches and the issues inherent in building up molecules from smaller fragment pieces. The ultimate goal of these calculations is to link two or more simulated fragments into a molecule that has an experimental binding affinity consistent with the additive predicted binding affinities of the virtual fragments. Computationally predicting binding affinities is a complex process, with many opportunities for introducing error. Therefore, care should be taken with the fragment preparation procedure to avoid introducing additional inaccuracies.

This chapter is focused on the preparation process used to create a virtual fragment database. Several key issues of fragment preparation which affect the accuracy of binding affinity predictions are discussed. The first issue is the selection of the two-dimensional atomic structure of the virtual fragment. Although the particular usage of the fragment can affect this choice (i.e., whether the fragment will be used for calibration, binding site characterization, hit identification, or lead optimization), general factors such as synthetic accessibility, size, and flexibility are major considerations in selecting the 2D structure. Other aspects of preparing the virtual fragments for simulation are the generation of three-dimensional conformations and the assignment of the associated atomic point charges.

**Key words** Fragment-based drug design, FBDD, Fragment screening, Virtual fragment, Fragment preparation, Conformations, Partial charges, Fragment database, Fragment linking

---

### 1 Introduction

The concept of using small molecule fragments in drug design is not new. The theoretical concept of linking fragments into molecules was presented by Jenks in 1981 [1]. In 1995, a review was published by Böhm regarding computational methods for joining fragments into molecules [2]. Fragment-based screening using NMR was described by Shuker et al. in 1996 [3]. In 2000, X-ray

crystallography was reported as another method for screening fragments by Nienaber et al. [4]. With experimental techniques, a fragment's binding to a protein can be detected in the high micromolar to millimolar range [3, 4]. In an increasing number of cases, screening fragments has led to successful hits where the traditional high-throughput screening of larger molecules did not [3–6]. A number of reviews report the progress of the fragment-based drug design (FBDD) field in greater detail [5–12].

There are multiple advantages to evaluating fragment–protein binding versus the binding of larger molecules. Synthesis timelines can be shortened, since a fragment-based approach is well suited to building molecules through parallel synthesis libraries. In-house chemistry developments can be leveraged in fragment-based design with proprietary fragments. Another advantage of using fragments is that fragment chemical space is much smaller than molecule chemical space; on the order of  $10^7$  for fragments with up to 11 heavy atoms [13] versus greater than  $10^{60}$  for molecules with up to 30 heavy atoms [14]. Hann and coworkers proposed that simpler molecules would have increased hit rates compared to those of complex molecules [15], and this theory has been verified experimentally. The Novartis group reported typical hit rates for fragment screens of 10–1,000 times that of traditional HTS [16]. The combination of a reduced chemical space to explore and higher hit rates means improved chemical diversity can be attained with molecules that are built up from fragments [7, 17]. These advantages increase the probability that a novel, potent inhibitor can be designed for a particular protein with the requisite physical properties for a drug molecule.

Fragment binding data can also facilitate the understanding of the protein binding site. It has been demonstrated that the ability of fragments to bind to an active site relates to the druggability of that site [18]. A fragment's binding location can give information about what chemotypes can have favorable interactions with the protein's binding site and what contribution those interactions have to whole-molecule binding, clarifying the structure-activity relationship (SAR) [12]. This knowledge can be used to shorten the time from hit to lead molecule and to increase the quality of the leads. Fragment–protein interactions can be observed that are excluded when the fragment is part of an elaborated molecule due to geometric, steric, or electrostatic constraints of the protein [15]. For that reason, observing a fragment's binding position may suggest scaffolds or connectivities that were not previously explored. Novel interactions that increase the potency of a molecule so that it can be truncated in other regions will create new intellectual property (IP) and potentially improve the selectivity and physical properties of the molecule. In the case of p38 MAP kinase, simulations of fragments with the allosteric X-ray structure led to the discovery of a new interaction with Arginine 70 on the  $\alpha$ C-helix [19]. A molecule was designed that utilized this novel interaction region

in the allosteric site and did not interact with the ATP hinge-region. This molecule was potent ( $IC_{50}=22$  nM) and selective against a panel of 150 kinases [19].

---

## 2 Computational Theory

Although there are advantages to experimental fragment-based screening over HTS, these methods also have drawbacks. As mentioned in the review by Konteatis, the high solubility necessary for the fragment screens and the experimental detection limits of fragment binding are disadvantages of the biophysical methods [12]. These techniques may also require large quantities of protein (0.5–5 mg) and have low experimental throughput [7]. In silico approaches are not bound by these constraints. By exploring fragment–protein binding with computational fragments, more possibilities can be evaluated. In theory, simulating two-thousand fragments represents a chemical space of  $8 \times 10^9$  three fragment molecules [17]. Virtual fragments hit rates, as well as the ability to connect these fragments geometrically and synthetically, generally leads to several hundred promising synthetic targets to evaluate experimentally [17]. Computational fragment methods (either alone or combined with experimental fragment screening) are contributing to the discovery of nanomolar inhibitors [5, 11, 19].

At Locus Pharmaceuticals two types of fragment–protein simulations were used: Grand Canonical Monte Carlo simulations where the excess chemical potential of the system is annealed, and systematic sampling that methodically and efficiently explores fragment binding positions [20–24]. For both methods the virtual fragments are treated as rigid bodies and the protein structure has a fixed conformation [22, 24]. The interaction energies between fragment and protein are calculated from non-bonded interaction parameters from the molecular mechanics force field AMBER [25], except for the fragment partial charges which are calculated with the quantum mechanics software Gaussian [26]. Since the fragments are regarded as rigid bodies, multiple conformations are generated for flexible fragments. Further molecule flexibility is accounted for when the fragments are linked to form molecules [27].

When linking the virtual fragments to form molecules, the binding energies of the fragments are assumed to be additive. The binding energy of the linked molecule is approximated by the sum of the fragment binding energies, even though this approximation does not account for the entropy loss upon linking the fragments [1, 8]. The preparation method of the virtual fragment also affects how well the assumption of binding energy additivity holds true. The selection of the fragment’s two-dimensional atomic structure, the generation of three-dimensional conformers, and the calculation of the fragment’s atomic point charges all affect the success of the fragment binding affinity predictions.

### 3 Methods for Preparing Computational Fragments

#### 3.1 Selection of the Two-Dimensional Atomic Structure

The first step in preparing a virtual fragment is to decide on the 2D atomic structure of that fragment. The fragment's 2D structure should be selected so that the properties of the fragment do not change significantly when that fragment is part of a molecule. There should not be a large change in the partial charge distribution or the geometry of the local energy minimum pose when the fragments are linked. The size of the fragment is also important. A fragment that is very small or does not have enough interacting groups may show non-selective binding to the protein (a desirable trait for non-interacting linkers). A large, complex fragment may generate an unmanageable number of conformations, or it may not bind to the protein due to steric constraints or charge repulsions. Based on observed fragment hits from X-ray crystallography screens, Jhoti and coworkers proposed a "Rule of Three" in 2003 [28] for building fragment databases. The "Rule of Three" guidelines (analogous to Lipinski's "Rule of Five" [29]) constrain a fragment's molecular weight to less than 300, with no more than 3 hydrogen bond donors or acceptors, and a Clog  $P \leq 3$ . The authors also suggest limits on rotatable bonds ( $\leq 3$ ) and polar surface area ( $\leq 60$ ) as possible selection criteria. In a 2013 follow-up to the "Rule of Three" proposal, Jhoti et al. state that the limitations on hydrogen bond donors and acceptors have not been generally implemented, due in part to ambiguities in how they should be defined [30]. There is experimental evidence that valuable fragments fall outside these guidelines [31] and this may be the case even more with *in silico* fragments.

Another consideration for selecting the 2D structure of a virtual fragment is if the synthetic intermediate for the fragment is commercially available or synthetically feasible. The chemistry to incorporate the fragment into a full molecule should also be reasonable. Ideally, the fragment should not contain any substructures that are unstable *in vivo* or known to produce toxicity in a drug. When selecting the 2D structure, the fragment should be evaluated for other stable enantiomers or charge states. Unless these states are generated by the simulation software, they should be prepared as distinct fragments for the database.

The intended purpose of the fragment may affect how the 2D atomic structure is selected. Fragments may be chosen as part of a diverse "probe" set, in order to map a protein's binding site and explore its affinity for different chemotypes. Other fragments may be proprietary, based on in-house synthetic schemes. Another method for generating fragments is to deconstruct known ligands into their fragment components. This approach is useful when reproducing known SARs and calibrating the simulation methodology for a particular protein, or when designing for families of proteins that may bind similar fragments. A large portion of the

Locus Fragment Database was generated by deconstructing compounds from medicinal chemistry literature, with emphasis on fragments found in known orally bioavailable drugs [21].

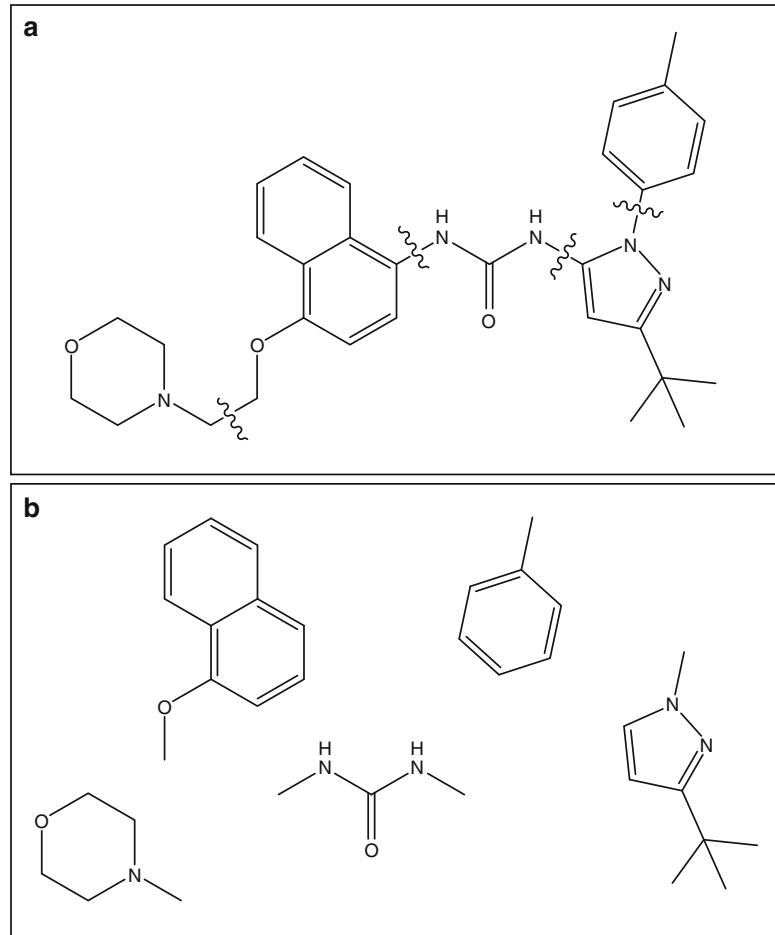
The 3D properties of the fragment also need to be considered when choosing the 2D structure. For instance, when an amide bond is added to an aromatic ring the electronic distribution of the system changes, since the aromaticity is maintained across the linkage. Therefore, more accurate results would be obtained by simulating a benzamide fragment instead of benzene and N-Methylacetamide. In other words, if the electronic distribution is different after connection, the binding of the separate fragments may be overestimated or underestimated compared to the linked entity. This adds additional inaccuracies to an already complex process. The differences between a fragment as a distinct entity versus as a substructure in a molecule should always be kept in mind when preparing fragments and building molecules. These issues cannot always be resolved through fragment selection, so the designed molecules should be further analyzed to verify that the desired protein ligand interactions are preserved [32].

A published p38 MAP kinase ligand [33] provides an example of deconstructing a ligand into the appropriate fragment components [12, 17]. If the molecule in question is the p38 kinase inhibitor BIRB-796 (Fig. 1a), a reasonable way to break up the fragments is as shown (Fig. 1b). As stated before, factors such as size, flexibility, and electronic distribution should be considered when selecting fragment substructures. Although one could simulate a single fragment containing the urea, *N*-methylpyrazole, and *t*-butyl moieties (Fig. 2), in the allosteric site of p38 the channel where the urea binds is narrow, with specific interactions to the urea. Only certain torsion angles around the urea–pyrazole bond will result in fragment geometries compatible with the constraints of the binding pocket. If other torsion angles are used, the compound fragment might be erroneously excluded from the pocket by the simulation. Hence, knowledge of the protein binding site can also influence fragment selection for a particular protein target.

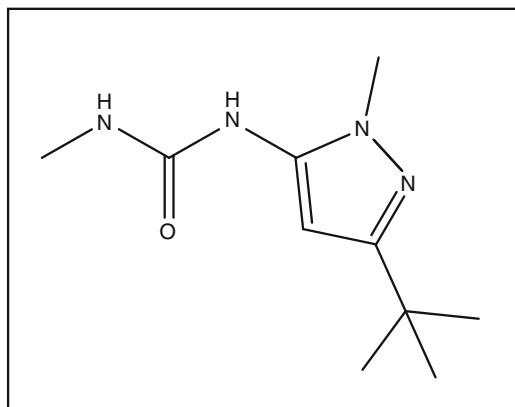
### 3.2 Conformer Generation

Some of the conformational information for molecules built by linking fragments is determined by where the fragments are positioned when they are joined [27]. However, since the simulations use rigid fragments, for a chemical structure that is actually flexible (i.e., has rotatable bonds or unsaturated ring systems), multiple 3D structures (i.e., conformers) must be prepared [21, 22]. The 2D structure of the fragment should be selected to minimize the number of rotatable bonds. Otherwise, the large number of conformers generated for the fragment will create an unmanageable number of simulations, with the associated data storage and data processing issues. The rule of thumb is to have three or less rotatable bonds in a fragment [28].

Changes in the geometry of the local energy minimum pose when the fragment becomes part of a molecule is also a concern



**Fig. 1** An example of deconstructing a known ligand into fragments. BIRB-796 is a potent binder to the p38 MAP Kinase allosteric site [13]. (a) BIRB-796 and the bonds to be broken to form fragments. (b) The 2D structures of the resulting fragments



**Fig. 2** Alternative fragment choice from BIRB-796

when preparing the 3D conformers of a fragment. This issue can be addressed by manually adding rotor states to the fragment conformer set that are known to be reasonable when the fragment is a substructure of a molecule. Another option is to simulate fragments with a steric placeholder, such as a methyl group, where the linkage would be. When a methyl placeholder is used, the software for linking fragments needs to handle removing the methyls or it will need to be done manually.

### 3.3 Partial Charge Assignment

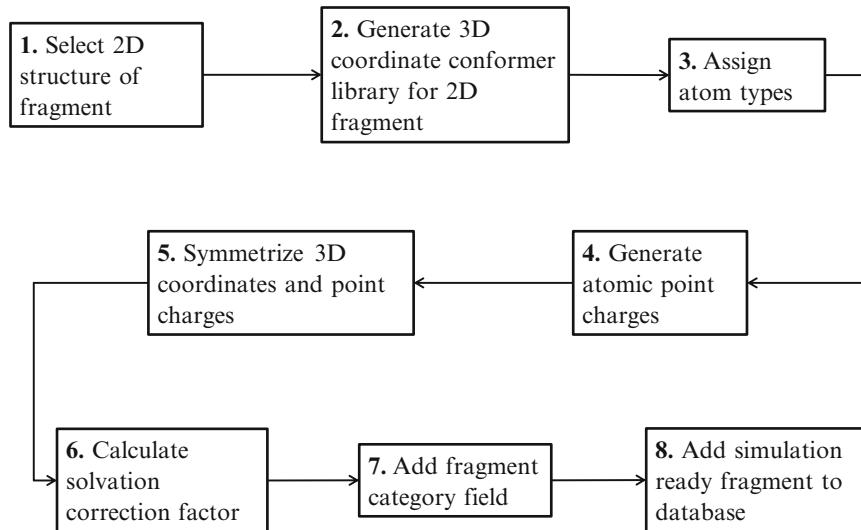
The partial (point) charges of a virtual fragment are another important component of binding affinity predictions. A study was done in 2003 comparing four methodologies to calculate geometries and charges of virtual fragments [34]. Since partial charges cannot be measured directly, solvation energies and dipole moments of 20 small fragments were calculated from the point charges and compared to experimental values. It was determined that the best compromise between speed and accuracy of the four methods was to use the molecular mechanics program MacroModel [35] with its implementation of the AMBER force field [36], AMBER\*, to generate conformations, and to use a Gaussian [26] CHelpG [37] single-point calculation with B3LYP [38–40] functionals and 6-31G(d) basis set [41] for the partial charges. The predicted binding affinity of ligands built from multiple fragments for three published protein systems were then evaluated with this method and compared to experimentally measured IC<sub>50</sub>s. The resulting standard deviation was  $\pm 1.0$  pIC<sub>50</sub> units/molecule [34]. Simulations of fragments prepared by the selected method were also used to reproduce X-ray crystal binding poses [19, 21, 22, 34], and to predict the binding affinities of designed compounds with a similar standard deviation to the literature compounds [34].

---

## 4 Fragment Preparation Workflow

The specific steps for preparing a virtual fragment for simulation are discussed below (Fig. 3). The fragment files and preparation techniques should be tailored to the requirements of the individual simulation software, therefore not all of the steps may be necessary.

1. The first step of preparing a fragment for simulation is to select the 2D structure according to the previously mentioned considerations.
2. After the fragment has been selected, the 2D representation is converted into a group of 3D conformers. Since the 2D structure of the fragment is selected to be small, with a limited number of rotatable bonds, the conformational analysis can be done by systematically rotating bonds and then energy minimizing the structure's coordinates. In this procedure, the



**Fig. 3** The work flow of the virtual fragment preparation process

systematic conformational search in MacroModel [35] is used to produce the fragment conformations. If the fragment is not flexible there will be only one conformation.

3. Since the fragment–protein simulations use a molecular mechanics force field, force field atom types are assigned to the fragment atoms. These atom types determine the force field parameters used for each atom during the simulations. For calculations with rigid fragments and a static protein, only the non-bonded parameters of the force field are needed. The force field used for fragments in this procedure is AMBER\*, the version of the AMBER force field [36] used in MacroModel [35]. The AMBER force field was originally developed for studying proteins and nucleic acids [36] but can also perform well when calculating protein–fragment interactions [19, 22, 24, 34].
4. The atomic point charges for each 3D conformation are calculated because the charge distributions are dependent upon the 3D structure. Since the geometry of the fragment is already determined, a single point calculation is performed in Gaussian [26]. The B3LYP [38–40] density functional method (which uses hybrid functionals) is used, along with the 6-31G(d) basis set [41]. CHelpG [37] is used to calculate the point charges from the electronic distribution.
5. Next, a symmetry operation is performed on the atomic coordinates and point charges. This accounts for rounding errors and maintains the planes of symmetry in the fragment.
6. The treatment of the solvent for fragment–protein simulations can be done by using a continuum dielectric, explicit water

molecules that are co-simulated, or a solvent correction factor that is based on the solvation energy of the fragment. This particular procedure was developed for preparing fragments that would be simulated in vacuo. Therefore, a solvent correction factor is calculated with the generalized Born surface area (GB/SA) method [42] as implemented in MacroModel [35].

7. A category field is created for the fragment. Labels like “Probe Set,” “Heterocyclic,” or “Kinase Hinge Pieces” can facilitate the extraction of fragments from the database. Different fragment sets may be used based on the stage of a project and the current design issues to be addressed.
8. Finally the fragment is stored in a database with a format that is compatible with the software used to perform the fragment–protein simulations.

---

## 5 Conclusion

Fragment-based drug design is emerging as a significant area of the drug discovery field. Fragment screening provides an efficient exploration of chemical space, and the ability to detect valuable hits that are missed with traditional screening methods. Computational approaches can avoid experimental limitations of solubility, protein supply, and throughput, but accuracy in the predicted binding affinities remains a concern. The preparation of the virtual fragments, including the selection of the 2D atomic structures and the generation of the 3D conformers and atomic partial charges, is critical to the success of the computational simulations.

---

## Acknowledgements

The author thanks the following colleagues for their contributions to the Locus technology and this approach to virtual fragment preparation: F. Guarneri, Z. Konteatis, T. Fujimoto, F. Hollinger, M. Clark, J. Wiseman, A. Klon, J. Zou, S. Meshkat, G. Talbot, K. Milligan, and W. Chiang. The author also thanks D. Ludington for editing assistance.

---

## References

1. Jencks WP (1981) On the attribution and additivity of binding energies. *Proc Natl Acad Sci U S A* 78:4046–4050
2. Böhm HJ (1995) Site-directed structure generation by fragment-joining. *Perspect Drug Discov Design* 3:21–33
3. Shuker SB, Hajduk PJ, Meadows RP et al (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274: 1531–1534
4. Nienaber VL, Richardson PL, Klighofer V et al (2000) Discovering novel ligands for

- macromolecules using X-ray crystallographic screening. *Nat Biotechnol* 18:1105–1108
5. Erlanson DA (2006) Fragment-based lead discovery: a chemical update. *Curr Opin Biotechnol* 17:643–652
  6. Murray CW, Verdonk ML, Rees DC (2012) Experiences in fragment-based drug discovery. *Trends Pharmacol Sci* 33:224–232
  7. Erlanson DA, McDowell RS, O'Brien T (2004) Fragment-based drug discovery. *J Med Chem* 47:3463–3482
  8. Rees DC, Congreve M, Murray CW et al (2004) Fragment-based lead discovery. *Nat Rev Drug Discov* 3:660–672
  9. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* 6:211–219
  10. Congreve M, Chessari G, Tisi D et al (2008) Recent developments in fragment-based drug discovery. *J Med Chem* 51:3661–3680
  11. Gozalbes R, Carbajo RJ, Pineda-Lucena A (2010) Contributions of computational chemistry and biophysical techniques to fragment-based drug discovery. *Curr Med Chem* 17:1769–1794
  12. Konteatis ZD (2010) In silico fragment-based drug design. *Expert Opin Drug Discov* 5:1047–1065
  13. Fink T, Bruggesser H, Reymond JL (2005) Virtual exploration of the small-molecule chemical universe below 160 daltons. *Angew Chem Int Ed* 44:1504–1508
  14. Bohacek RS, McMullan C, Guida WC (1996) The art and practice of structure-based drug design: a molecular modeling perspective. *Med Res Rev* 16:3–50
  15. Hann MM, Leach AR, Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci* 41:856–864
  16. Schuffenhauer A, Ruedisser S, Marzinzik A et al (2005) Library design for fragment based screening. *Curr Top Med Chem* 751–762
  17. Klon AE, Konteatis Z, Meshkat SN et al (2011) Fragment and protein simulation methods in fragment based drug design. *Drug Dev Res* 72:130–137
  18. Hajduk PJ, Huth JR, Fesik SW (2005) Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* 48:2518–2525
  19. Moffet K, Konteatis Z, Nguyen D et al (2011) Discovery of a novel class of non-ATP site DFG-out state p38 inhibitors utilizing computationally assisted virtual fragment-based drug design (vFBDD). *Bioorg Med Chem Lett* 21:7155–7165
  20. Guarnieri F, Mezei M (1996) Simulated annealing of chemical potential: a general procedure for locating bound waters. Application to the study of the differential hydration propensities of the major and minor grooves of DNA. *J Am Chem Soc* 118:8493–8494
  21. Moore WR (2005) Maximizing discovery efficiency with a computationally driven fragment approach. *Curr Opin Drug Discov Devel* 8:355–364
  22. Clark M, Guarnieri F, Shkurko I et al (2006) Grand canonical Monte Carlo simulation of ligand-protein binding. *J Chem Inf Model* 46:231–242
  23. Clark M, Meshkat S, Wiseman J (2009) Grand canonical free-energy calculation of protein-ligand binding. *J Chem Inf Model* 49: 934–943
  24. Clark M, Meshkat S, Talbot GT et al (2009) Fragment-based computation of binding free energies by systematic sampling. *J Chem Inf Model* 49:1901–1913
  25. Cornell WD, Cieplak P, Bayly CI et al (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197
  26. Frisch MJ, Trucks GW, Schlegel HB et al (1998) Gaussian 98 (revision A9). Gaussian, Inc., Pittsburgh, PA
  27. Clark M, Meshkat S, Talbot G et al (2009) Developing technologies in biodefense research: computational drug design. *Drug Dev Res* 70:279–287
  28. Congreve M, Carr R, Murray C et al (2003) A ‘rule of three’ for fragment-based lead discovery? *Drug Discov Today* 8:876–877
  29. Lipinski CA, Lombardo F, Dominy BW et al (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23:3–25
  30. Jhoti H, Williams G, Rees DC et al (2013) The ‘rule of three’ for fragment-based drug discovery: where are we now? *Nat Rev Drug Discov* 12:644
  31. Köster H, Craan T, Brass S et al (2011) A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J Med Chem* 54:7784–7796
  32. Konteatis ZD, Klon AE, Zou J et al (2011) Computational approach to de novo discovery of fragment binding for novel protein states. *Methods Enzymol* 493:357–380

33. Pargellis C, Tong L, Churchill L et al (2002) Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *J Nat Struct Biol* 9:268–272
34. Ludington JL, Fujimoto TT, Hollinger FP (2004) Determining partial atomic charges for fragments used in de novo drug design. 228th ACS national meeting, Philadelphia, PA (Poster)
35. Mohamadi F, Richard NGJ, Guida WC et al (1990) Macromodel – an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J Comput Chem* 11:440–467
36. Weiner SJ, Kollman PA, Case DA et al (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc* 106:765–784
37. Breneman CM, Wiberg KB (1990) Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J Comput Chem* 11: 361–373
38. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A* 38:3098
39. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 37:785
40. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98:5648–5652
41. Hariharan PC, Pople JA (1973) The influence of polarization functions on molecular orbital hydrogenation energies. *Theor Chim Acta* 28: 213–222
42. Qiu D, Shenkin PS, Hollinger FP et al (1997) The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J Phys Chem A* 101:3005–3014

# Chapter 5

## Fragment Library Design: Using Cheminformatics and Expert Chemists to Fill Gaps in Existing Fragment Libraries

Peter S. Kutchukian, Sung-Sau So, Christian Fischer, and Chris L. Waller

### Abstract

Fragment based screening (FBS) has emerged as a mainstream lead discovery strategy in academia, biotechnology start-ups, and large pharma. As a prerequisite of FBS, a structurally diverse library of fragments is desirable in order to identify chemical matter that will interact with the range of diverse target classes that are prosecuted in contemporary screening campaigns. In addition, it is also desirable to offer synthetically amenable starting points to increase the probability of a successful fragment evolution through medicinal chemistry. Herein we describe a method to identify biologically relevant chemical substructures that are missing from an existing fragment library (chemical gaps), and organize these chemical gaps hierarchically so that medicinal chemists can efficiently navigate the prioritized chemical space and subsequently select purchasable fragments for inclusion in an enhanced fragment library.

**Key words** Fragment based screening, Cheminformatics, Chemogenomics, Library design, Privileged substructure, Chemical space, Chemical diversity

---

### 1 Introduction

The design of fragment libraries has garnered much attention as fragment based lead discovery has transitioned from an exploratory field for lead identification to a more standard industrial strategy to identify novel starting points for hit to lead campaigns. The composition of a fragment library may be based on the assay technology that is employed—for example fluorine containing fragments for NMR [1], bromine containing fragments for X-ray crystallography [2], or fragments amenable to SPR [3]. On the other hand, a library may be designed for general use for multiple assay technologies [4]. In addition, libraries may be designed to include biologically active motifs [5]. Furthermore, the input of chemists on the desirability of fragments for optimization has been incorporated into the design of fragment libraries [6].

Many research groups have a standard fragment library which at some point, will need to be updated for a variety of reasons. Primary among these is to replace compounds that have historically behaved badly (e.g., promiscuity, inactivity) or to replenish compounds that are no longer available. Herein we describe a simple procedure that can easily be carried out in an academic or industrial setting to identify “chemical gaps”—i.e., desirable substructures that are present in biologically active compounds but not in the current fragment library—and identify purchasable compounds that fill these gaps. The chemical gaps are organized in a hierarchical fashion so that chemists can easily traverse chemical space in a Spotfire [7] session, and select purchasable fragments to fill the gaps.

---

## 2 Materials

### 2.1 Filling Chemical Gaps

1. Carry out all computational procedures with Pipeline Pilot version 9.1 (Bovia) [8] unless otherwise specified (*see Note 1*).
2. Commercially available catalogues of fragments are obtained directly from vendors in a structure format (e.g., SD file) or in a text file that includes SMILES that are converted to structures in Pipeline Pilot (*see Note 2*).
3. We use Merck’s Chemical Genetic Interaction Enterprise (CHEMGENIE) Database for chemogenomic data (*see Note 3*). CHEMGENIE is comprised of both internal dose-response data such as PanLabs, Invitrogen Kinase Screen, and High-Throughput Screening and external dose-response data such as ChEMBL (EMBL-EBI) [9, 10] and Metabase (Thomson Reuters) [11] as well as high quality biophysical binding data such as ALIS [12].
4. Visualization of chemical gaps is performed in Spotfire version 5.5.1 (TIBCO) [7].

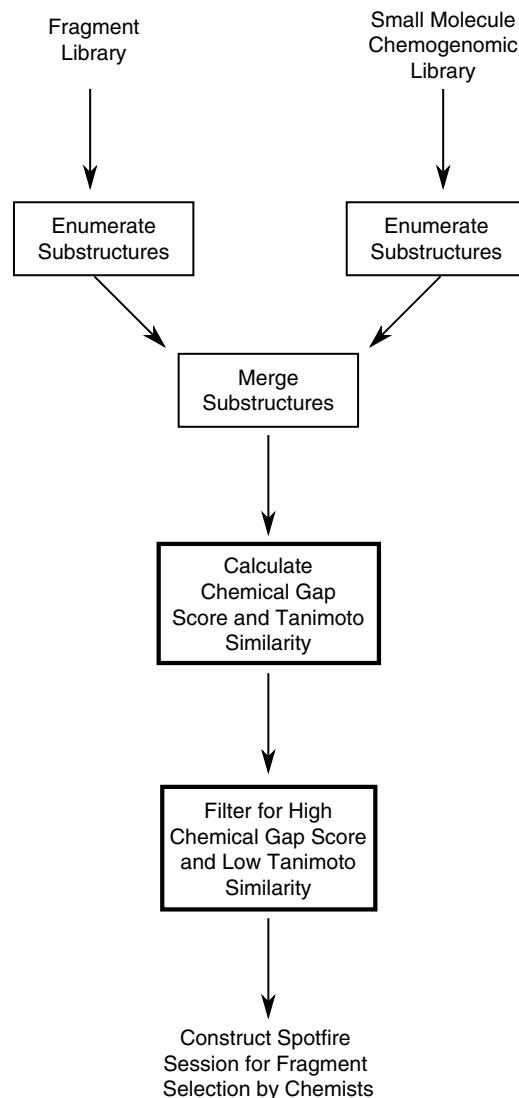
---

## 3 Methods

### 3.1 Filling Chemical Gaps

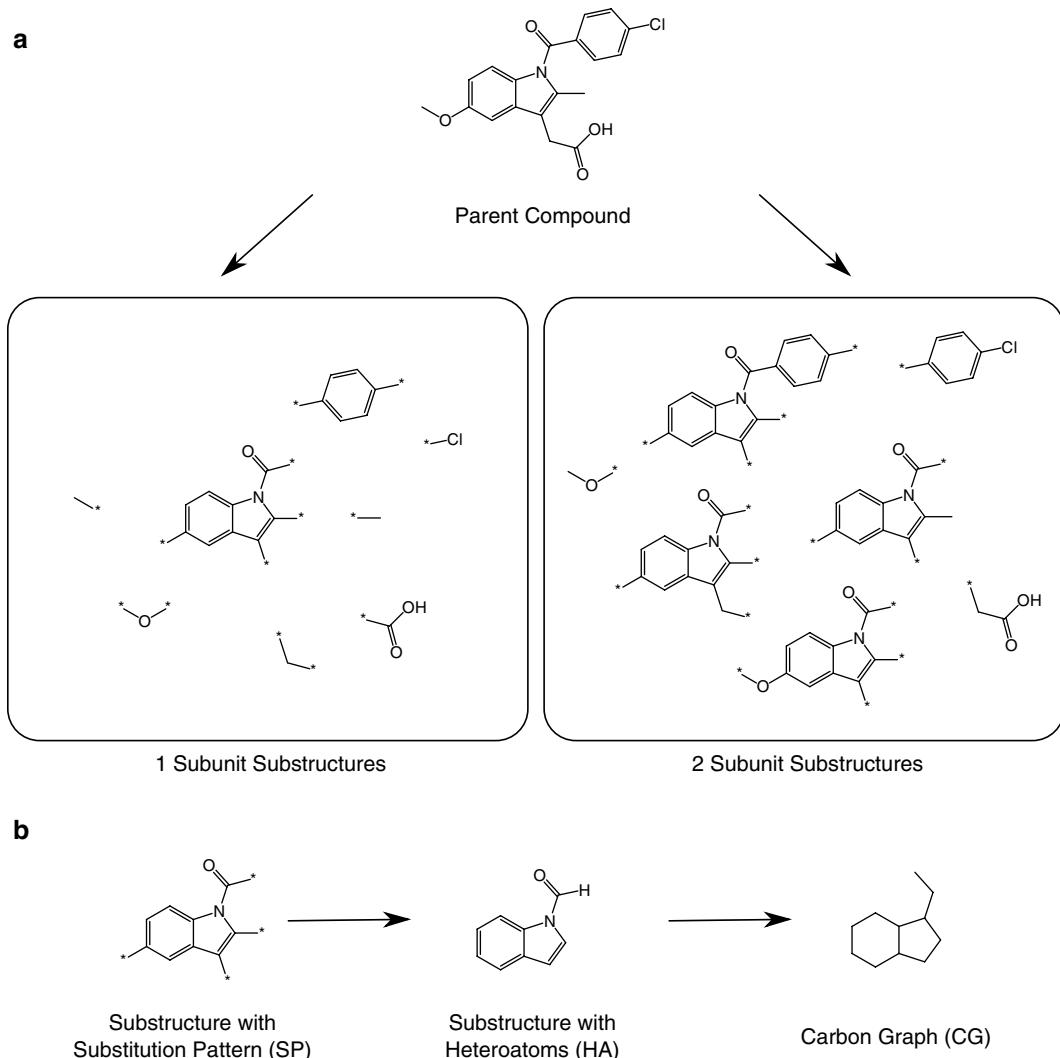
A high level flow diagram for the method is depicted in Fig. 1.

1. A Pipeline Pilot protocol is used to convert the current fragment library and chemogenomic library (CHEMGENIE) to hierarchical sets of SMILES and corresponding InChIKeys (Fig. 2, *see Note 4*). The three sets of SMILES/InChIKeys are carbon graph substructures (CG, with some aromatic information—*see Note 5*), heteroatom substructures (HA), and substitution pattern substructures (SP).
2. First de-salt and standardize all compounds in Pipeline Pilot.



**Fig. 1** High level flow diagram of method. All steps enclosed by boxes are performed in Pipeline Pilot

3. Fragment each compound into one subunit substructures (Fig. 2a, see Note 6). Convert each resulting substructure with substitution pattern (SP) into a substructure with hetero-atoms (HA), and a carbon graph (CG) as in Fig. 2b. To do this, use the “Map Substructure” component twice—once using a SMARTS file of disallowed bond breaks, and once using a SMARTS file with allowed bond breaks. Then use the “Enumerate Fragments” component, specifying the “allowed bond breaks” and “disallowed bond break” properties as the “Property with Bonds to Break” and “Property with Bonds to Skip” parameters, respectively. Add “Mg” as “Atoms to Add



**Fig. 2** Fragmentation of compounds. (a) A parent compound is fragmented into 1 subunit or 2 subunit substructures using bond break and bond protect rules. (b) A substructure with substitution pattern (SP) is converted to a substructure with heteroatoms (HA) but no substitution pattern, and finally to a carbon graph (CG) substructure

at Break Points". Prior to visualizing the SP SMILES string, however, convert Mg atoms to [\*] using the "Reaction from SMIRKS" component, and a SMIRKS string that converts Mg to [\*]. To generate the HA substructures, convert all Mg to H atoms. To generate the CG substructure, use the "Standardize Molecule" component to convert all non-hydrogen atoms to carbon and to make all bonds single.

4. Count each type of substructure (CG, HA, and SP) in three separate data streams, and merge the counts obtained from the fragment library and from CHEMGENIE, so that there are three separate tables written to file with the following

properties: substructure SMILES, substructure InChIKey, count in fragment library (Library A), count in CHEMGENIE Bioactives (Library B). Also, the InChIKey for the substructure that is one hierarchical step towards being more general should be written out. Thus, for SP table, the HA InChIKey should be written out, and for the CG table the HA InChIKey should be written out. These will be used to relate tables in spotfire. Thus, there will be a SP, HA, and CG table.

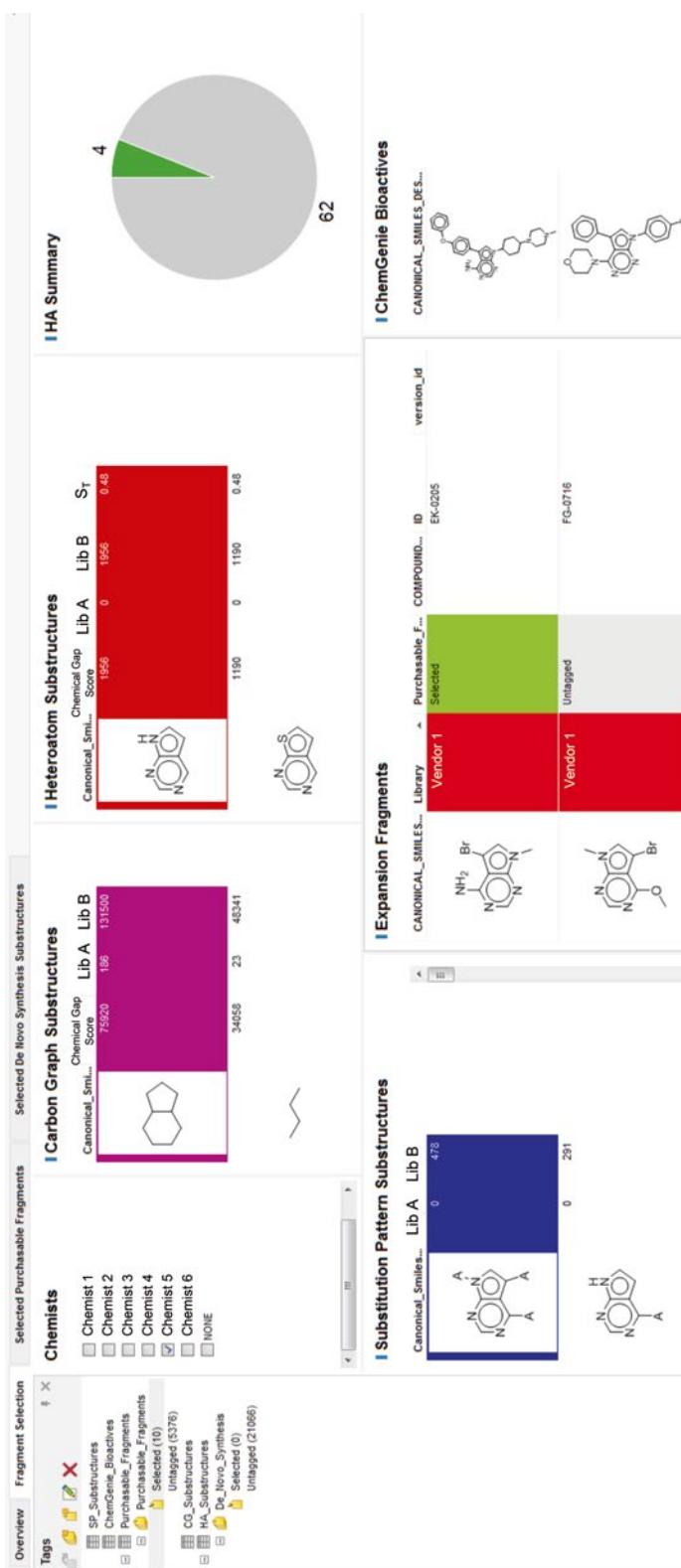
5. To prioritize filling chemical gaps—substructures that are present in bioactive CHEMGENIE compounds (library B), but not in the fragment library (library A)—calculate a Chemical Gap Score for the CG substructures and for the HA substructures (*see Note 7*). The chemical gap score for a substructure  $i$  is the sum of occurrences of substructure  $j$  in library B (CHEMGENIE Bioactives) if the substructure  $j$  is not present in library A (fragments):

$$\text{Chemical Gap Score}(i) = \sum_j \begin{cases} N_j^B & \text{if } N_j^A = 0 \\ 0 & \text{else} \end{cases}$$

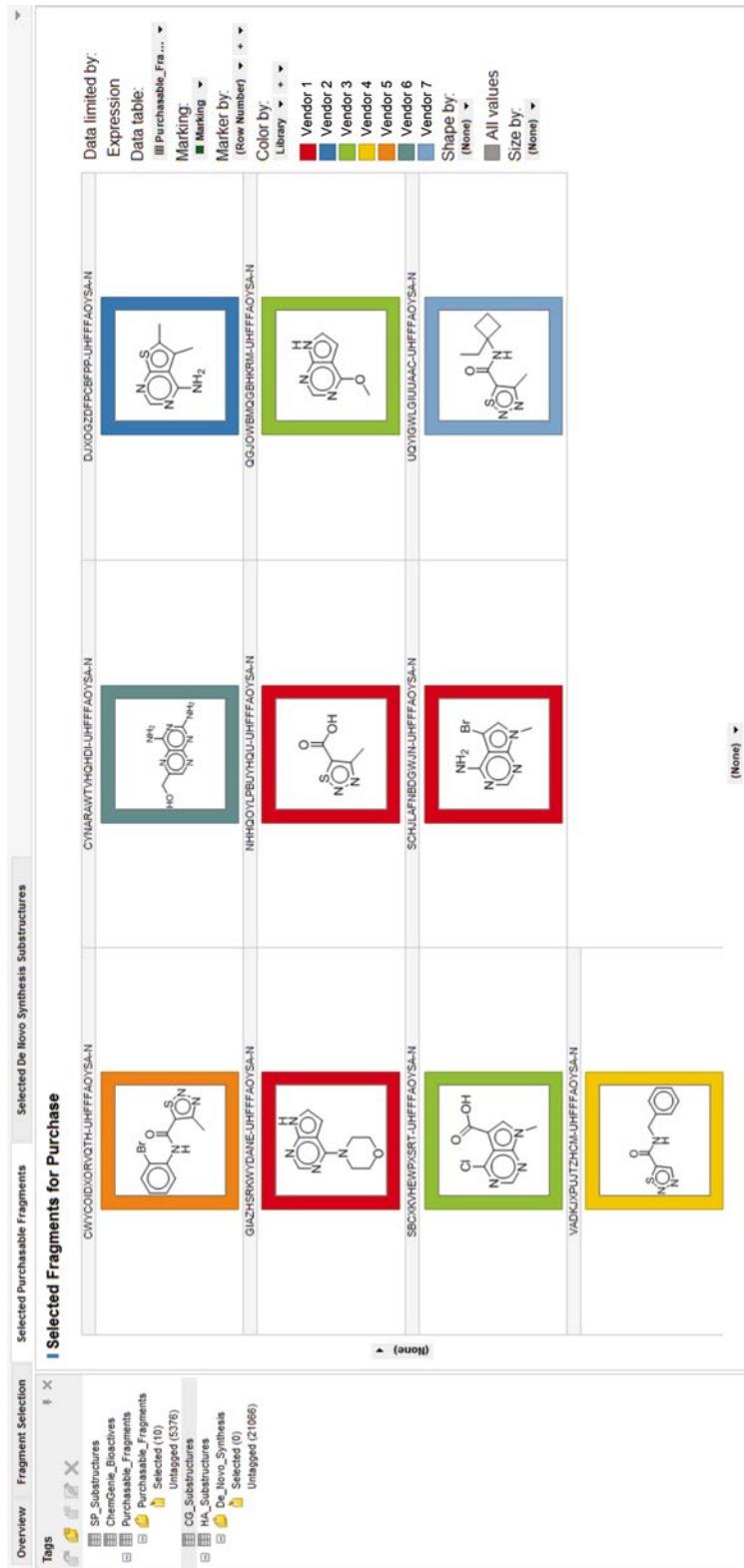
where  $i$  is hierarchically related to  $j$  (e.g., if  $i$  corresponds to a CG substructure, then  $j$  corresponds to all HA substructures that share that carbon graph, or if  $i$  corresponds to a HA substructure, then  $j$  corresponds to all SP substructures that share that HA substructure),  $N_j^A$  is the occurrences of substructure  $j$  in library A, and  $N_j^B$  is the occurrences of substructure  $j$  in library B. Note, to prioritize SP substructures, the occurrence of the SP substructure in the CHEMGENIE Bioactives is used.

6. Prioritize HA substructures based on similarity to HA substructures in current fragment library. Calculate the maximum nearest neighbor Tanimoto similarity ( $S_T$ ) to HA substructures present in the current fragment library. HA substructures with low  $S_T$  are prioritized (*see Note 8*).
7. Use chemical gap score *and* similarity to select high priority substructures that are missing from the current fragment library. Filtering is performed on the HA substructure level. All HA substructures with  $S_T < 0.5$  and a chemical gap score  $>100$  are selected (*see Note 9*). Filter the HA, and CG substructure tables so that they only include substructures that will link to the high priority substructures. Only high priority HA substructures are visualized in Spotfire.
8. Divide the CG substructure that correspond to high priority HA substructures equally among chemists that will be evaluating purchasable fragments (*see Note 10*). In this manner, one chemist will evaluate similar HA substructures (e.g., similar ring systems with different substitution patterns), and can select a desirable number of neighbors from the purchasable fragments.

9. Identify fragments that can be purchased for the purpose of filling chemical gaps.
10. For each compound in the commercial fragment library, identify all SP substructures and HA substructures. Generate a table that includes the SP substructure InChIKey, HA substructure InChIKey, parent compound identifier, parent compound SMILES, and source.
11. Filter the previous table (**step 10**) for substructures (HA) that are high priority. That is, read in all high priority substructures (HA) identified above (**step 7**) to an array, and only include substructures in the purchasable set of fragments that are contained in this array. This will be the purchasable fragments table in spotfire.
12. Prepare the CHEMGENIE bioactive compounds for inclusion in the spotfire selection session. Fragment and filter the compounds as in **steps 10** and **11**. This will be the CHEMGENIE Bioactives table in spotfire.
13. Assemble spotfire session for purchasable fragment selection as follows (*see Note 11*). First, import tables and add relationships. Import the carbon graph (CG) substructure table, the hetero-atom (HA) substructure table, the substitution pattern (SP) substructure table, the purchasable fragments table, and the CHEMGENIE Bioactives table. Add table relationships to relate the tables to each other in the following manner, where the column used to relate the table is in parenthesis. The CG and HA table (CG InChIKey), the HA and SP table (HA InChIKey), the SP and purchasable fragments table (SP InChIKey), the SP and CHEMGENIE Bioactives table (SP InChIKey).
14. Add a filter on the CG table so that the chemist selecting fragments can be selected (Fig. 3).
15. Add CG substructures, HA substructures, SP substructure, purchasable fragments, and CHEMGENIE bioactives table visualizations to spotfire session (Fig. 3).
16. Add “Purchasable\_Fragments” tag collection to Purchasable\_Fragments table, and add “Selected” tag.
17. Each chemist initiates a selection session by selecting their name. A CG is then selected, which populates the HA substructures. An HA substructure is then selected, which populates the SP substructures. An SP substructure is then selected, which populates purchasable fragments, as well as CHEMGENIE Bioactive compounds that contain the SP substructure that is missing from the current fragment library (Fig. 3).
18. Chemists can then “tag” the purchasable fragments as “Purchasable\_Fragments:Selected.”
19. All tagged fragments can be reviewed (Fig. 4). Fragments can also be untagged in this tab.



**Fig. 3** Screenshot of a spotfire session used by chemists to select purchasable fragments that fill chemical gaps. First, a chemist selects one's name. Then the CG substructure, HA substructure, and SP substructure are selected. This populates the purchasable fragments and CHEMGENIE Bioactives (for context). Purchasable fragments can then be tagged as selected. The HA Summary window serves as a summary for the total number of purchasable fragments corresponding to an HA substructure, and how many have been tagged (green)



**Fig. 4** Screenshot of tagged purchasable fragments. The fragments are colored by the library source

---

## 4 Notes

1. Protocols with sample input files are available upon request.
2. It is assumed that the commercially available fragments libraries are filtered for desirable physical chemical properties (e.g., the rule of 3) [13] and substructures (e.g., to remove reactive or pan assay interference compounds) [14] prior to inclusion in the following method.
3. In this procedure, we use the chemogenomics database CHEMGENIE to identify biologically relevant substructures missing from the current fragment library. Another strategy, however, might entail using a company's current high-throughput screening (HTS) deck as a library source, fragmenting the deck, and counting the occurrence of substructures (or biologically active substructures). This alternate strategy ensures good overlap between the fragment library and the HTS library, and increases the probability that virtual fragment linking (VFL) could be applied successfully to expand from fragments to small molecules [15, 16].
4. A chemogenomic database is required to associate chemical structures with their biological activity. Ideally, the chemogenomic database should include data that has been integrated from internal and external repositories.
5. For the carbon graph InChIKey, we sought to capture some of the aromatic information in the ring system as well. Thus, we used a modified key which was the carbon graph InChIKey concatenated with the number of ring bonds and number of aromatic bonds originally present prior to converting all bonds to single bonds. For example, an indole key would be BNRNAKTVFSZAFA-UHFFFAOYSA-N\_10\_10.
6. In the current implementation, we use specific bond break/protect rules to fragment compounds into substructures. These rules are intended to obtain substructures that are meaningful to medicinal chemists (i.e., no rings are broken in half, and functional groups are protected). However, there are alternative strategies to derive substructures, for example, retrosynthetic transformations can be employed to derive substructures with synthetic handles [17].
7. In this procedure we use the count of a particular substructure in the CHEMGENIE Bioactives to prioritize the substructures that were missing in the current fragment library. There are other ways of measuring the biological importance of substructures, however. For example, chemogenomics data like CHEMGENIE could be used to assess the count of the number of target classes that a substructure has been present in, or

whether the substructure is enriched for a certain number of targets or target classes based on a Bayesian score for the substructure [18].

8. We filter away HA substructures with a high Tanimoto similarity to existing HA substructures in the fragment library, to focus chemists' selections on novel HA substructures present in purchasable fragments. This approach, however, is prone to remove HA substructures that exist in the fragment library, although the library might not contain fragments that encompass all substitution patterns present in the CHEMGENIE Bioactives for a given HA substructure. This concern could be addressed by removing the Tanimoto Similarity filter, and focusing on substructures that are missing at the SP level.
9. These filters can be modified to increase or decrease the potential set of chemical gaps and purchasable fragments to assess.
10. There are alternative strategies to divide the chemical gaps and corresponding purchasable compounds among chemists. For example, divisions could take place at the HA or SP substructure level.
11. In this procedure, we build a Spotfire session that allows chemists to tag fragments for purchase. If there are chemical gaps that do not correspond to purchasable fragments, however, these gaps could serve as inspiration for de novo fragment synthesis designs. In this case, the HA substructure could be tagged as a template for de novo synthesis.

## References

1. Dalvit C (2009) NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discov Today* 14:1051–1057
2. Blaney J, Nienaber V, Burley SK (2006) Fragment-based Lead Discovery and Optimization Using X-Ray Crystallography, Computational Chemistry, and High-throughput Organic Synthesis. In: Jahnke W, Erlanson DA (eds) Fragment-based approaches in drug discovery. Wiley, Weinheim, pp 215–248
3. Elinder M, Geitmann M, Gossas T et al (2011) Experimental validation of a fragment library for lead discovery using SPR biosensor technology. *J Biomol Screen* 16:15–25
4. Lau WF, Withka JM, Hepworth D et al (2011) Design of a multi-purpose fragment screening library using molecular complexity and orthogonal diversity metrics. *J Comput Aided Mol Des* 25:621–636
5. Mok NY, Brenk R, Brown N (2014) Increasing the coverage of medicinal chemistry-relevant space in commercial fragments screening. *J Chem Inf Model* 54:79–85
6. Kutchukian PS, Vasilyeva NY, Xu J et al (2012) Inside the mind of a medicinal chemist: the role of human bias in compound prioritization during drug discovery. *PLoS One* 7:e48476
7. TIBCO Software Inc (2014) TIBCO Spotfire, 5.5.1
8. Biovia (2013) Pipeline Pilot, 9.1
9. Overington JP (2009) ChEMBL: large-scale mapping of medicinal chemistry and pharmacology data to genomes. *Abstr Pap Am Chem S* 238
10. Bento AP, Gaulton A, Hersey A et al (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res* 42:D1083–D1090
11. Metabase. <http://thomsonreuters.com/metabase/>

12. Annis DA, Nickbarg E, Yang X et al (2007) Affinity selection-mass spectrometry screening techniques for small molecule drug discovery. *Curr Opin Chem Biol* 11:518–526
13. Congreve M, Carr R, Murray C et al (2003) A rule of three for fragment-based lead discovery? *Drug Discov Today* 8:876–877
14. Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* 53:2719–2740
15. Crisman TJ, Bender A, Milik M et al (2008) "Virtual fragment linking": an approach to identify potent binders from low affinity fragment hits. *J Med Chem* 51:2481–2491
16. Wassermann AM, Kutchukian PS, Lounkine E et al (2013) Efficient search of chemical space: navigating from fragments to structurally diverse chemotypes. *J Med Chem* 56:8879–8891
17. Lewell XQ, Judd DB, Watson SP et al (1998) RECAP – retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J Chem Inf Comp Sci* 38:511–522
18. Lounkine E, Kutchukian P, Glick M (2013) Chemometric applications of naïve Bayesian models in drug discovery – beyond compound ranking. In: Bajorath J (ed) *Chemoinformatics: case studies and pharmaceutical applications*. Wiley, Hoboken, NJ

## **Part II**

### **Simulation**

# Chapter 6

## Protocol for Fragment Hopping

Kevin B. Teuscher and Haitao Ji

### Abstract

Fragment hopping is a fragment-based approach to designing biologically active small molecules. The key of this approach is the determination of the minimal pharmacophoric elements in the three-dimensional space. Based on the derived minimal pharmacophoric elements, new fragments with different chemotypes can be generated and positioned to the active site of the target protein. Herein, we detail a protocol for performing fragment hopping. This approach can not only explore a wide chemical space to produce new ligands with novel scaffolds but also characterize and utilize the delicate differences in the active sites between isofunctional proteins to produce new ligands with high target selectivity/specifity.

**Key words** Fragment-based drug discovery, Fragment hopping, Scaffold diversity, Isofunctional proteins, Protein–protein interactions, Inhibitors, Selectivity, Peptidomimetics

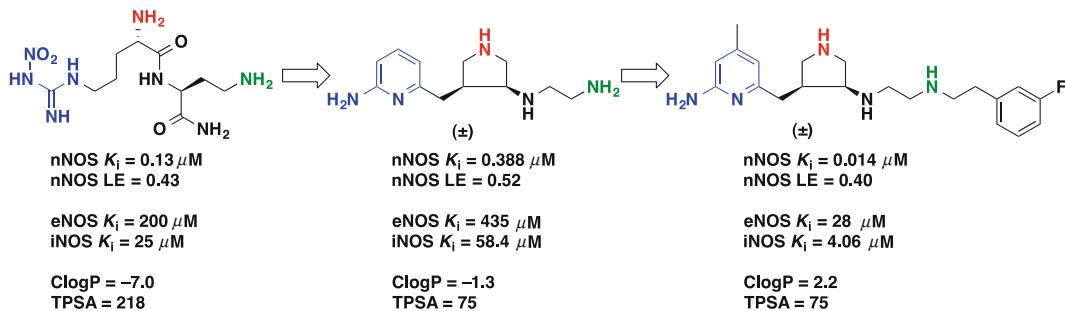
---

### 1 Introduction

Fragment-based drug design is widely used as an effective tool for lead discovery [1, 2]. The starting point for fragment-based drug design is the identification of low molecular weight fragments that bind to a target of interest by fragment screening. Biophysics-based techniques such as nuclear magnetic resonance (NMR) [3–7], X-ray crystallography [8–11], mass spectrometry (MS) [12, 13], surface plasmon resonance (SPR) spectroscopy [14–16], and confocal fluorescence correlation spectroscopy [17, 18] have been applied to screen fragment libraries. After fragment hits are identified, fragment evolution, fragment linking, and in situ fragment assembly (including dynamic combinatorial chemistry [19–21], tethering with extenders [22, 23], and in situ click chemistry [24–26]) have been employed to convert fragment(s) into a ligand molecule and maintain drug-like properties of the generated ligand. There are internal limitations and problems for the drug discovery projects that are initiated by fragment screening. First, these approaches only cover a small fraction of the total

chemical diversity space. It is estimated that a collection of  $10^3$  fragments can typically sample the chemical diversity space of  $10^9$  molecules. Although the combinatorial advantage of fragment screening makes a significant increase relative to high throughput screening (HTS), it is still a small fraction of the total chemical diversity space ( $10^{60}$ ). Second, in fragment screening fragment molecules can bind to the protein binding sites that are different from the binding sites of lead-like molecules [27–29]. Third, fragment screening only identifies and characterizes fragments for potency, that is, the fragments that bind to energetic hotspots (the regions on protein surface that are the major contributors to ligand free energy of binding). In fact, many binding sites that are responsible for target specificity and/or selectivity are not included in hotspots. The identification of fragments that bind to the sites responsible for target specificity is a crucial step for the discovery of selective inhibitors. There is a great desire to design small-molecule inhibitors for a specific target while leaving other related targets unaffected. Based on the factor that the structures of the biologically active compounds for a specific target are discontinuous points in chemical space, scaffold hopping has been used to identify compounds that have similar biological activities but totally different scaffolds [30, 31]. These methods can decrease the risks of construction for bioactive molecules, increase hit rates, and offer structural diversity. However, in scaffold hopping the skeleton of the newly designed molecules is confined to the basic architecture of the template structure, which usually comes from a known drug or drug candidate. In addition, mimicking the template structure using different scaffolds often does not optimize ligand–protein interactions to the maximal extent, because the biologically relevant groups in the template structure do not offer an optimal match between a small molecule and the biological target.

We proposed fragment hopping to initiate the design of potent and selective small-molecule inhibitors [32]. Fragment hopping is a pharmacophore-driven strategy. The core of this approach is the derivation of the minimal pharmacophoric elements for key binding pockets. Then, fragments with different chemotypes are generated to match the requirement of minimal pharmacophoric elements. Therefore, fragment hopping can explore a wider chemical space. Fragment hopping determines the positioning of the minimal pharmacophoric elements and then places fragments to match the generated minimal pharmacophoric elements, leading to a higher success rate for fragment identification. After fragment linking or evolving, the pharmacophores in the new molecule maintain the same spatial orientation as those in the minimal pharmacophoric element model. Fragment hopping



**Fig. 1** Fragment hopping to design highly potent and selective inhibitors for neuronal nitric oxide synthase

determines the minimal pharmacophoric elements that are important for ligand selectivity and generates fragments to trigger inhibitor selectivity, providing an efficient pathway to generate selective inhibitors. Compared with scaffold hopping, fragment hopping not only maps important interaction patterns between a ligand and the protein target based on a priori scaffold but also makes it possible to generate more diverse scaffolds. Furthermore, fragment hopping takes into account synthetic accessibility in fragment linking and evolving through the use of the bioisosteric replacement technique.

Since fragment hopping was proposed in 2008 and applied to design highly potent and selective inhibitors for neuronal nitric oxide synthase (nNOS) as shown in Fig. 1 [32, 33], a few successful examples using the fragment hopping concept to design small-molecule inhibitors have been reported for the other enzymes [34, 35] and protein–protein interaction targets [36, 37]. The key step of all these reported studies was the determination of the pharmacophores. However, the procedures to generate lead-like molecules have been diverse, and a different efficiency for lead generation has been observed for different approaches. There is a need to overview current studies and outline a general procedure for a productive fragment hopping design.

## 2 Materials

Table 1 lists common programs that can be used in fragment hopping. Table 2 indicates the fragment libraries that are used in fragment hopping.

**Table 1**  
**Common programs used in fragment hopping**

Program	Version	Provider	URL
GRID <sup>a</sup>	22c	Molecular Discovery Ltd.	<a href="http://www.moldiscovery.com">www.moldiscovery.com</a>
SiteMap	2.6	Schrödinger	<a href="http://www.schrodinger.com">www.schrodinger.com</a>
MCSS <sup>a</sup>	2.1	Accelrys/Discovery Studio 4.0	<a href="http://www.accelrys.com">www.accelrys.com</a>
GOLPE <sup>a</sup>	4.5	Multivariate Infometric Analysis S.r.l.	<a href="http://www.miasrl.com/golpe.htm">www.miasrl.com/golpe.htm</a>
FTMap <sup>a</sup>		Professor Sandor Vajda	<a href="http://www.ftmap.bu.edu">www.ftmap.bu.edu</a>
Computational alanine scanning <sup>a</sup>		Accelrys/Discovery Studio	<a href="http://www.accelrys.com">www.accelrys.com</a>
ANCHOR <sup>a</sup>		Professor Carlos J. Camacho	<a href="http://www.structure.pitt.edu/anchor">www.structure.pitt.edu/anchor</a>
PocketQuery <sup>a</sup>		Professor Carlos J. Camacho	<a href="http://www.pocketquery.csb.pitt.edu">www.pocketquery.csb.pitt.edu</a>
Catalyst <sup>b</sup>		Accelrys/Discovery Studio 4.0	<a href="http://www.accelrys.com">www.accelrys.com</a>
GALAHAD <sup>b</sup>		Certara/Tripos	<a href="http://www.tripos.com/index.php">www.tripos.com/index.php</a>
Phase <sup>b</sup>	3.4	Schrödinger	<a href="http://www.schrodinger.com">www.schrodinger.com</a>
LigandScout <sup>b</sup>	3.1	Inte:Ligand	<a href="http://www.inteligand.com">www.inteligand.com</a>
MOE pharmacophore <sup>b</sup>		MOE	<a href="http://www.chemcomp.com">www.chemcomp.com</a>
CoMFA, COMSIA <sup>b</sup>		Certara/Tripos	<a href="http://www.tripos.com/index.php">www.tripos.com/index.php</a>
LUDI		Accelrys/Discovery Studio 4.0	<a href="http://www.accelrys.com">www.accelrys.com</a>
SEED		Professor Amedeo Caflisch	<a href="http://www.biochem-caflisch.uzh.ch/download/">www.biochem-caflisch.uzh.ch/download/</a>
GLIDE	5.8	Schrödinger	<a href="http://www.schrodinger.com">www.schrodinger.com</a>
DOCK	3.7	Professor Brian K. Schoichet	<a href="http://www.dock.compbio.ucsf.edu/DOCK3.7/">www.dock.compbio.ucsf.edu/DOCK3.7/</a>
FFLD		Professor Amedeo Caflisch	<a href="http://www.biochem-caflisch.uzh.ch/download/">www.biochem-caflisch.uzh.ch/download/</a>
AutoDock	4.2	Professor Arthur J. Olson	<a href="http://www.autodock.scripps.edu">www.autodock.scripps.edu</a>
Cscore		Certara/Tripos	<a href="http://www.tripos.com/index.php">www.tripos.com/index.php</a>
MetaSite	4.1	Molecular Discovery Ltd.	<a href="http://www.moldiscovery.com">www.moldiscovery.com</a>
DAIM <sup>c</sup>		Professor Amedeo Caflisch	<a href="http://www.biochem-caflisch.uzh.ch/download/">www.biochem-caflisch.uzh.ch/download/</a>

<sup>a</sup>Used to derive the pharmacophores based on the structure of the protein target

<sup>b</sup>Used to derive the pharmacophores based on the structures of a set of known bioactive compounds or the ligand peptides/proteins

<sup>c</sup>Used to deconstruct known drugs and drug candidates

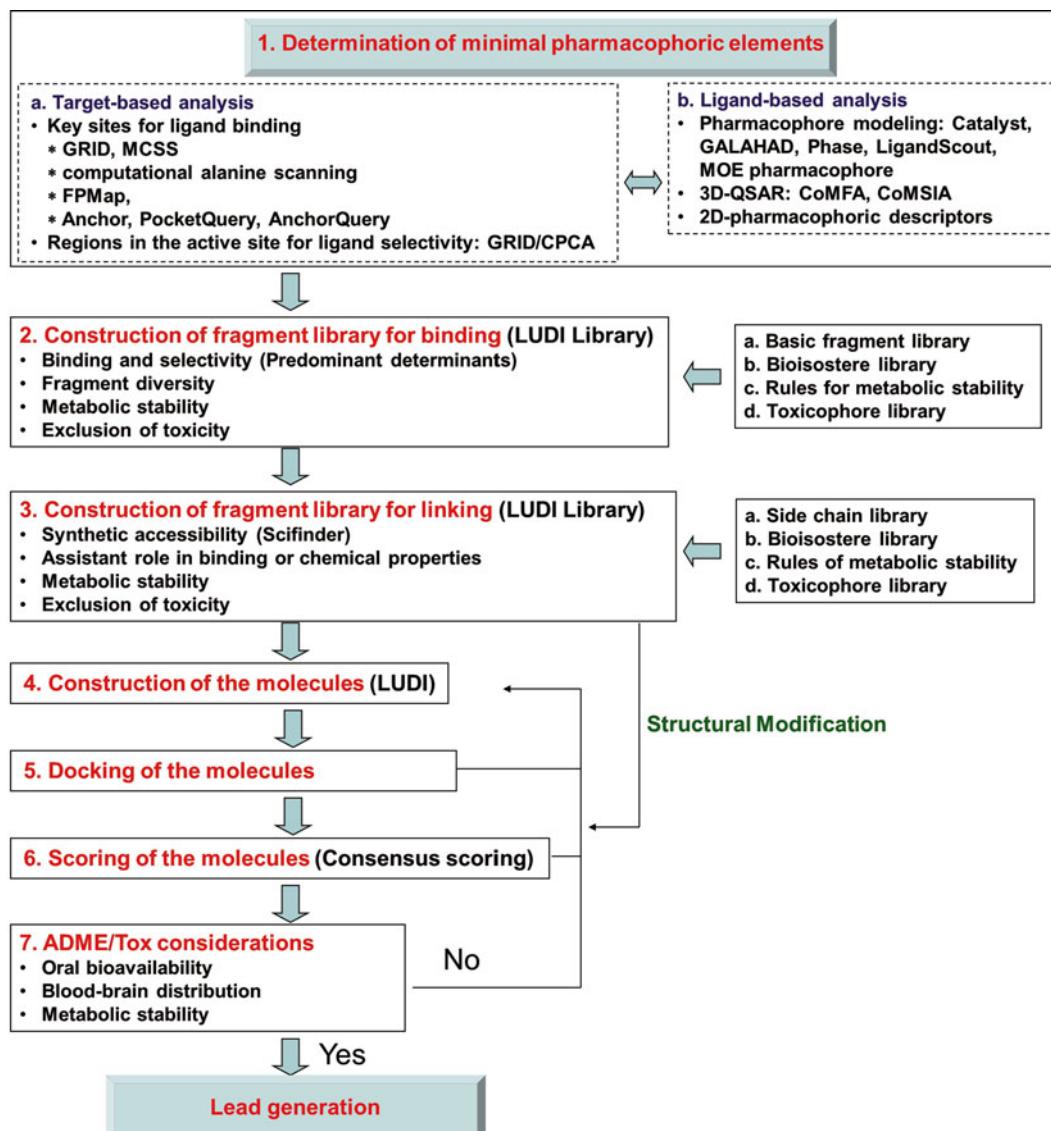
**Table 2**  
**Databases used in fragment hopping**

Name	Source
Basic fragment library	<i>ACS Chem Biol</i> 2013, 8, 524–529, Supplementary Figure 5
Bioisostere library	<i>ACS Chem Biol</i> 2013, 8, 524–529, Supplementary Figure 4
Rules for metabolic stability	<i>J Am Chem Soc</i> 2008, 130, 3900–3914, Supplementary Figure 3
Toxicophore library	<i>J Am Chem Soc</i> 2008, 130, 3900–3914, Supplementary Figure 4
Side chain library	<i>ACS Chem Biol</i> 2013, 8, 524–529, Supplementary Figure 6

### 3 Methods

A schematic flow diagram for fragment hopping is shown in Fig. 2 (*see Note 1*).

- When the protein target is an enzyme and the structure of the enzyme target is known, GRID [38, 39], SiteMap [40], and multiple copy simultaneous search (MCSS) [41, 42] in Table 1 are used to reveal the key binding sites for inhibitor potency.
  - Go to RCSB protein data bank (<http://www.rcsb.org/pdb/>) to download the protein structure, identify the ligand binding site, and define the regions of interest. Typically the residues that are within 6 Å from the center of the regions of interest were included in the calculations (*see Note 2*).
  - GRID was used to calculate 3D energy maps around the ligand binding site and highlight the favorable sites for a specific functional group of an organic compound. Hydrogen atoms were added by program GRIN. The GRID box dimensions were chosen to encompass all of the active site residues. The grid spacing was typically set to 1 Å. Directive NPLA (number of planes of GRID points per Angstrom) was set to 1. The amino acids in the ligand binding site were considered rigid (directive move is set to 0) or flexible (directive move is set to 1). The other settings were the standard default parameters. The typical probes used in the GRID modeling were DRY, C3, NM3, N1+, N3+, N1, NH=, O, O1, COO<sup>-</sup>, amidine, and ARamidine. The DRY probe was used to calculate the hydrophobic interactions. The C3 and NM3 probes were used to describe steric interactions. N1+, N3+, NM3, and COO<sup>-</sup> were used to explore charge–charge interactions. The others probes were used to determine the sites for hydrogen-bonding and other electrostatic interactions.



**Fig. 2** Schematic flow diagram for fragment hopping, a pharmacophore-driven strategy to generate selective inhibitors. Reprinted with permission from *J. Am. Chem. Soc.* (Ji, H., Stanton, B. Z., Igarashi, J., Li, H., Martásek, P., Roman, L.J., Poulos, T.L., Silverman, R.B. (2008) Minimal pharmacophoric elements and fragment hopping, an approach directed at molecular diversity and isozyme selectivity. Design of selective neuronal nitric oxide synthase inhibitors. *J Am Chem Soc* 130, 3900–3914). Copyright 2008 American Chemical Society

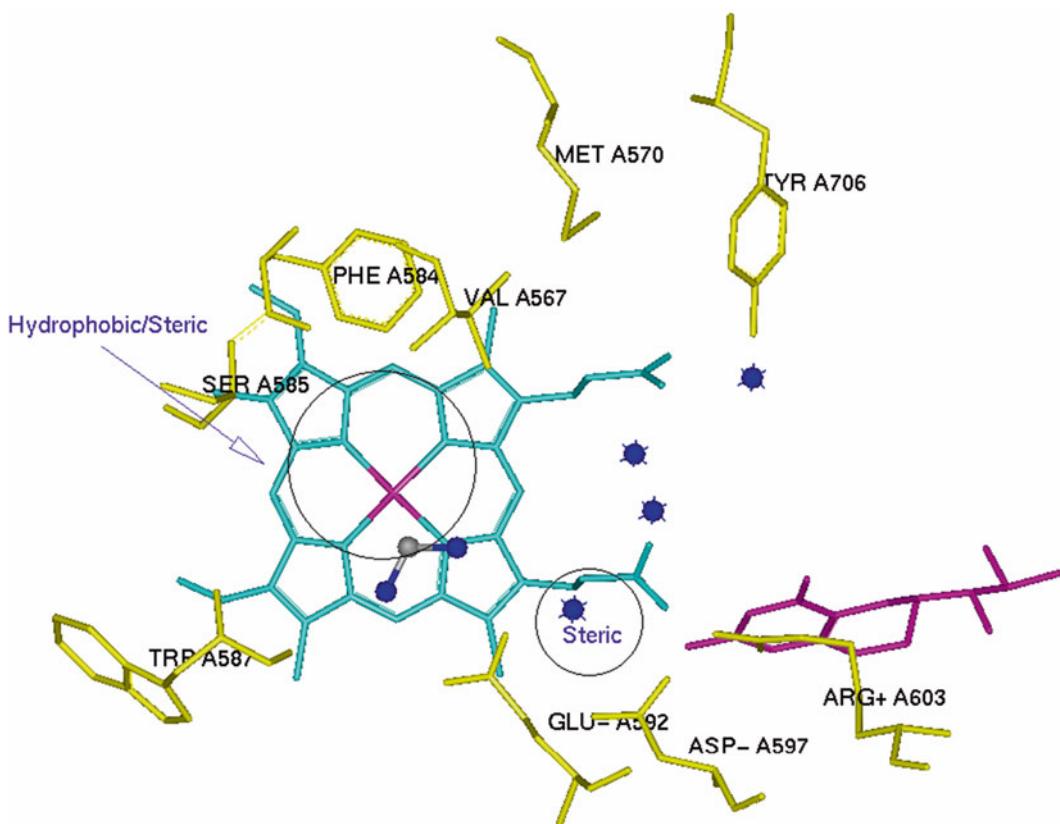
(c) MCSS randomly places thousands of copies of small functional groups into the ligand binding site, and the copies of small functional groups are subjected to energy minimization. The copies with the lowest energies highlight the potential regions for ligand binding. Hydrogen atoms were added to protein structure by Accelrys/Discovery Studio.

The CHARMM force field and the Momany–Rone partial charge [43] were applied to minimize the orientation of hydrogen atoms. About 2,500–5,000 replicas of a given functional group were randomly distributed in the defined binding site and then simultaneously and independently energy minimized. Pairs of molecules were considered identical if the root-mean-square deviation (rmsd) between them was  $<0.2$  Å. In such cases, one of the pairs with a lower binding energy was eliminated. The distance threshold for generating random fragments was set to 0.9 Å. The other settings were the standard default parameters. The MCSS calculations were performed using the CHARMM22 force field and the MCSS 2.1 program. Common functional groups used in the MCSS calculations are benzene, cyclohexane, propane, isobutane, *N*-methylacetamide, methanol, ether, acetate ion, methylammonium, and trimethylammonium.

2. When the protein target is an enzyme and the structures of the isoenzymes are known, GRID/consensus principal component analysis (CPCA) can be used to map the key binding site for inhibitor selectivity [44, 45]. The molecular interaction fields (MIFs) from the GRID calculations were imported into the GOLPE program, as shown in Table 1. A maximum cutoff was set to 0 kcal/mol to consider only the favorable protein–ligand interactions (negative energy values). The positive interaction energy is in most cases due to unfavorable steric repulsions between the probes and the atoms in the box. Because the equations used for calculating the MIF values are very different for different probes, block unscaled weights (BUW) were used to normalize the interaction energies between different probes so that each probe would get the same importance in the model. Variables with the values  $<0.01$  kcal/mol and those with a standard deviation below 0.02 were removed to eliminate noisy variables. The pretreated data were then used in CPCA modeling. CPCA is a hierarchical principal component analysis (PCA) and can capture the information in both the blocks (i.e., individual probes) and the whole X-matrix (i.e., MIFs values). Therefore, CPCA can be regarded as a PCA at two different levels: one is the block level, which provides the relative importance of different probes; another is the superlevel, which is the combination of these blocks to yield an analysis for the overall data (the results are similar to that obtained from the usual PCA). Each of these levels has loading and score vectors that summarize the information like in a usual PCA. In a selectivity study, often more than one principal component contributes to discriminate different objects in the scores plot; therefore, any single CPCA loadings plot can only partially describe the MIFs difference for a specific GRID probe between different target proteins. By using active

CPCA differential plots implemented in the GOLPE program, the difference between the two points for the first and second principal components can be calculated and projected back into the original space (a pseudofield) using PCA loading. That is, the vector linking pairs of objects in a two-dimensional scores plot can be translated into isocontour plots that identify those variables that contribute most to differentiating selected objects. Two stages of the GRID/CPCA analyses are typically made. One includes all residues of the ligand binding site. The second one targets a limited region for a specific subpocket using the cutout tool of GOLPE.

3. When the target is a protein–protein interface, FTMap [46, 47], computational alanine scanning [48], AnchorQuery [49], and PocketQuery [50, 51] in Table 1 can be used to derive the key sites for protein–protein binding and small-molecule inhibition (*see Note 3*).
  - (a) If the structure of the target protein is known, Web-based FTMap can be used to predict the energetic hotspots for ligand potency. The PPI mode of the FTMap was selected to ensure a better prediction for the shallow pockets on the protein surface.
  - (b) If the structure of the protein–protein complex is known, computational alanine scanning can be used to predict key residues that significantly contribute to the protein–protein interaction. Click “Design Protein” from the “Macromolecule” menu of Accelrys Discovery Studio and then “Calculate Mutation Energy (Binding)”. Assign “Mutate To” alanine and “Implicit Solvent Dielectric Constant” to 80 to derive key residues for protein–protein interactions.
  - (c) If the structure of the protein–protein complex is known, Web-based ANCHOR [52] can be used to identify the protruding hotspot residues whose solvent accessible surface area (SASA) displays a  $>0.5 \text{ \AA}^2$  change upon protein–protein complexation. PocketQuery can be used to predict concave hot regions. Both searches simply require an upload of the structure for the protein–protein complex. The output data are the changes in SASA ( $\Delta\text{SASA}$ ) and the FastContact energy ( $\Delta G$ ). PocketQuery also provides the Rosetta energy ( $\Delta\Delta G$ ) that equals to the  $\Delta\Delta G$  from computational alanine scanning, the residue conservation scores, and a consensus druggability score for the concave hot region.
4. Establishment of minimal pharmacophoric elements. The minimal pharmacophoric elements can be an atom, a cluster of atoms, a virtual graph, or vectors. It is an ensemble of electronic and steric features that is essential for binding to a specific



**Fig. 3** Minimal pharmacophoric elements for selective nNOS inhibitors. An amidino group is positioned close to E592. A yellow nitrogen atom is close to D597. The regions where hydrophobic and/or steric interactions play important roles are indicated by circles. Three blue nitrogen atoms are placed close to the heme propionate. Reprinted with permission from *J. Am. Chem. Soc.* (Ji, H., Stanton, B. Z., Igarashi, J., Li, H., Martásek, P., Roman, L. J., Poulos, T. L., Silverman, R. B. (2008) Minimal pharmacophoric elements and fragment hopping, an approach directed at molecular diversity and isozyme selectivity. Design of selective neuronal nitric oxide synthase inhibitors. *J Am Chem Soc* 130, 3900–3914). Copyright 2008 American Chemical Society

pocket in the ligand binding site. Figure 3 shows a minimal pharmacophore element model for selective nNOS inhibitors (see Notes 4 and 5).

5. The basic fragment library in Table 2 was manually examined to find all of the possible fragments that can match the requirements of the minimal pharmacophore elements for each pocket. The bioisostere library was manually examined to generate a focused fragment library with diverse structures. The generated focused fragment library was then interrogated with the rules for metabolic stability and a toxicophore library shown in Table 2 to provide potential fragments for a specific pharmacophore.
6. The targeted fragment libraries were converted into a LUDI user library [53, 54]. To determine the appropriate position and orientation of the fragments from the LUDI user library,

LUDI was first applied to generate the interaction sites for each pharmacophore. Four different types of the interaction sites are defined in the LUDI program: lipophilic–aliphatic, lipophilic–aromatic, hydrogen bond donor, and hydrogen bond acceptor. The residues inside an 8.0 Å radius sphere, which centered on the centroid of the minimal pharmacophoric elements, were used to generate the interaction sites. The link library switch was turned off, and the target mode switch was turned on. The LUDI scoring function was set to Energy-Estimate–2. The other settings were kept with the standard default parameters (*see Note 6*).

7. The generated fragments were then docked to the ligand binding site using Schrödinger Glide XP and to relax the flexible bonds in the fragment molecules [55, 56]. All fragments were charged by OPLS\_2005. For fragment docking, any atoms within 6 Å from the proposed critical pharmacophoric elements were used to define the grid box. The number of poses per ligand for the initial docking stage was set to 50,000. A wider scoring window of 500.0 kcal/mol was used to keep initial poses, and the best 1,000 poses per ligand were kept for energy minimization. The keyword *roughmin* was added to the *maxkeep* line of the Glide input file to instruct Glide to bypass sorting by the rough score and to minimize all *maxkeep* poses on the Glide grid. The maximum number of minimization steps was set to 5,000. One hundred poses were subjected to post-docking minimization, and 100 poses were recorded for output. The other settings were the default Glide XP parameters. A visual inspection was performed to examine the degree of match between the docked fragment poses and the proposed critical pharmacophoric elements (*see Note 7*).
8. To convert the newly generated fragments into a molecule, the side chain library in Table 2 was converted into a LUDI link library and used for the connection operation. The hydrogen atoms in the above fragment structures were replaced by a link fragment to create a new substructure. The LUDI switch for the target mode was turned off, but the LUDI switch for the link library was turned on. The linkage parameter can be set to 1 (the link fragment fits at least one link site), 2 (the link fragment simultaneously fits at least two link sites), or was specified (the link site was specifically assigned) according to the actual requirements. The other settings were the standard default parameters. The bioisostere library in Table 2 and the SciFinder® search are two useful tools for generating synthetically feasible molecules.
9. AutoDock 4.2 [57] was used to dock lead-like molecules to the protein target. Hydrogen atoms were added to the ligand molecules and the partial atomic charges were calculated using

the Gasteiger–Marsili method [58]. The protonation state of the target protein was set to pH 7.0 when adding hydrogen atoms. The partial atomic charges of the target protein were calculated by AMBER 7 FF99. The AMBER 7 force field 99 within SYBYL X2.0 was used to optimize the orientation of the hydrogen atoms on the target protein. In AutoDock 4.2 calculations, the grid maps were calculated using AutoGrid with the grid spacing of 0.375 Å. For the ligand docking study, the dimensions of the grid box included all residues of the active site. Docking was performed using the Lamarckian genetic algorithm (LGA), and the pseudo-Solis and Wets method was applied for the local search. Each docking experiment was performed 100 times, yielding 100 docked conformations. Parameters for the docking experiments were as follows: the number of individuals in the population was 150; the maximum number of energy evaluation was 2,500,000; the maximum number of generations was 27,000; the number of top individuals to survive to the next generation was 1; the rate of gene mutation was 0.02; the rate of crossover was 0.8; the iterations of Solis and Wets local search was 300; the local search rate was 0.06. The other settings were the standard default parameters. The auxiliary clustering analysis was then used to evaluate the results of the docking experiments.

10. The CScore module of SYBYL X2.0 was further used to evaluate the binding mode of the lead-like molecules. The AutoDock Binding Energy and the LUDI binding score were retrieved from the AutoDock and LUDI studies, respectively. The G\_Score, PMF\_Score, D\_Score, and ChemScore were calculated using the Tripos CScore module. Open a spreadsheet table that contains the ligand molecules. Then, create the ORIGIN attribute by typing “table attribute create ORIGIN string”, and set it to DOCKING by typing “table attribute set ORIGIN DOCKING”. Save the spreadsheet table and reopen it. Run Cscore to obtain G\_Score, PMF\_Score, D\_Score, and ChemScore. AutoDock, and LUDI scores can be imported manually. Click Consensus to generate the consensus score for each docked molecule. Two other scoring functions commonly used for consensus scoring were ASP [59] and ChemPLP [60].
11. The last step is ligand-based prediction to evaluate absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox) of the designed ligand molecule. Lipinski’s rule of five (H-bond donors  $\leq 5$ , H-bond acceptors  $\leq 10$ , molecular weight  $\leq 500$ , and  $\log P \leq 5$ ) [61] and the polar surface area (PSA)  $\leq 140 \text{ \AA}^2$  [62] was used to predict oral bioavailability of the designed molecule. The blood–brain barrier penetration was predicted by the following rules of thumb [63, 64]: the number of nitrogen and oxygen atoms  $\leq 5$ ,  $\text{Clog } P - (\text{N+O}) > 0$ , PSA  $< 90 \text{ \AA}^2$ , molecular

weight  $\leq 450$ , and the logarithm of distribution coefficient of a compound ( $\log D$ ) is between 1 and 3. Rule of four (molecular weight  $> 400$ ,  $A \log P > 4$ , rings  $> 4$ , and H-bond acceptors  $> 4$ ) has been used to define the profile of a protein–protein interaction inhibitor [65]. The metabolically labile sites were predicted using MetaSite 4.1 [66]. The conformations of the ligand molecule were generated by molecular dynamics simulated annealing. The system was heated at 1,000 K for 1.0 ps and then annealed to 250 K for 1.5 ps. The annealing function was exponential. One hundred such cycles of annealing were performed and the resulting 100 conformers were subjected to energy minimization. Energy minimization was performed using the Tripos force field, Powell optimization method, and MAXIMIN2 minimizer with a convergence criterion of 0.05 kcal/mol·Å. Charges were calculated using the Gasteiger–Marsili method [58]. The conformations of the ligand molecules were clustered based on the rmsd value of non-hydrogen atoms  $< 1.0$ . The site of metabolism prediction module of MetaSite 4.1 was used to predict the metabolically labile sites of the ligand molecules (see Note 8).

---

## 4 Notes

1. The results of the experimental approaches are useful for deriving the minimal pharmacophoric elements. When the target enzyme structure is known, the results of the site-directed mutagenesis and fragment screening using NMR, X-ray crystallography, MS-based tethering, SPR, and substrate activity screening (SAS) [67, 68] are valuable for the determination of the pharmacophores. When the structure of the protein–protein interaction target is known, alanine scanning can be performed to identify and quantify the hotspots for protein–protein interactions.
2. The correct input protein structure is essential for almost all of the computer modeling programs. Therefore, it would be beneficial to clean the protein structure by removing the alternate conformations, patching missing side chains, and adding hydrogens before it is subjected to computer modeling. The orientations of hydrogens need be energetically optimized.
3. Neither GRID nor MCSS has the desolvation term in the scoring potential, but FTMap does. FTMap is more appropriate for mapping the energetic hotspots for protein–protein interactions.
4. When a set of bioactive small-molecule inhibitors is obtained for a specific target, ligand-based pharmacophore mapping is also important for deriving the minimal pharmacophoric elements. The commonly used pharmacophore modeling programs include Catalyst [69–71], GALAHAD [72, 73], Phase

[74, 75], LigandScout [76], and MOE pharmacophore. The structure–activity relationship analysis, the 3D-QSAR analysis such as CoMFA [77], CoMSIA [78], and the GRID/GOLPE analysis [79], and the 2D-pharmacophoric descriptors [80] are also useful for deriving the minimal pharmacophoric elements for both inhibitor potency and inhibitor selectivity. When the crystal structure of the protein–protein complex is known, the geometries and electronic properties of the projecting hotspots from the ligand peptides/proteins should be derived and used to determine the minimal pharmacophoric elements.

5. The mapping of the minimal pharmacophoric elements is user-dependent and iterative. If the result of ligand design from the first round is unsatisfactory, it is necessary to redefine the pharmacophore model. Based on our experience, the experimental results for protein–ligand binding are dependent upon efficient mapping of the minimal pharmacophoric elements.
6. LUDI has primarily been used to place the fragment into the corresponding binding site to match the proposed minimal pharmacophoric elements. SEED [81, 82] can be used as an alternative.
7. Glide XP has primarily been used to dock the generated fragments to match the proposed minimal pharmacophoric elements. DOCK [83], GOLD [84], and fragment-based flexible ligand docking (FFLD) [85] can be used as alternatives.
8. In fragment hopping, the protein structures are treated rigid, which is a drawback if protein flexibility needs to be considered in the design of new inhibitors with new scaffolds.

---

## Acknowledgement

This work was supported by the Pulmonary Fibrosis Foundation I.M. Rosenzweig Young Investigator Award (Award number: 235170) and the Department of Defense CDMRP BCRP breakthrough award (W81XWH-13-BCRP-BREAKTHROUGH).

## References

1. Murray CW, Rees DC (2009) The rise of fragment-based drug discovery. *Nat Chem* 1: 187–192
2. Davis BJ, Erlanson DA (2013) Learning from our mistakes: the ‘unknown knowns’ in fragment screening. *Bioorg Med Chem Lett* 23: 2844–2852
3. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274: 1531–1534
4. Hajduk PJ, Gerfin T, Boehlen JM, Häberli M, Marek D, Fesik SW (1999) High-throughput nuclear magnetic resonance-based screening. *J Med Chem* 42:2315–2317
5. Peng JW, Lepre CA, Fejzo J, Abdul-Manan N, Moore JM (2002) Nuclear magnetic resonance-based approaches for lead generation in drug discovery. *Methods Enzymol* 338:202–230
6. Mayer M, Meyer B (2001) Group epitope mapping by saturation transfer difference

- NMR to identify segments of a ligand in direct contact with a protein receptor. *J Am Chem Soc* 123:6108–6117
7. Dalvit C, Fogliatto G, Stewart A, Veronesi M, Stockman B (2001) WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J Biomol NMR* 21: 349–359
  8. Mattos C, Ringe D (1996) Locating and characterizing binding sites on proteins. *Nat Biotechnol* 14:595–599
  9. Lesuisse D, Lange G, Deprez P, Bénard D, Schoot B, Delettre G, Marquette JP, Broto P, Jean-Baptiste V, Bichet P, Sarubbi E, Mandine E (2002) SAR and X-ray. A new approach combining fragment-based screening and rational drug design: application to the discovery of nanomolar inhibitors of Src SH2. *J Med Chem* 45:2379–2387
  10. Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H (2005) Fragment-based lead discovery using X-ray crystallography. *J Med Chem* 48:403–413
  11. Ciulli A, Williams G, Smith AG, Blundell TL, Abell C (2006) Probing hot spots at protein-ligand binding sites: a fragment-based approach using biophysical methods. *J Med Chem* 49: 4992–5000
  12. Erlanson DA, Braisted AC, Raphael DR, Randal M, Stroud RM, Gordon EM, Wells JA (2000) Site-directed ligand discovery. *Proc Natl Acad Sci U S A* 97:9367–9372
  13. Seth PP, Miyaji A, Jefferson EA, Sannes-Lowery KA, Osgood SA, Propst SS, Ranken R, Massire C, Sampath R, Ecker DJ, Swayze EE, Griffey RH (2005) SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. *J Med Chem* 48:7099–7102
  14. Nordström H, Gossas T, Hämäläinen M, Källblad P, Nyström S, Wallberg H, Danielson UH (2008) Identification of MMP-12 inhibitors by using biosensor-based screening of a fragment library. *J Med Chem* 51:3449–3459
  15. de Kloet GE, Retra K, Geitmann M, Källblad P, Nahar T, van Elk R, Smit AB, van Muijlwijk-Koezen JE, Leurs R, Irth H, Danielson UH, de Esch IJP (2010) Surface plasmon resonance biosensor based fragment screening using acetylcholine binding protein identifies ligand efficiency hot spots (LE hot spots) by deconstruction of nicotinic acetylcholine receptor  $\alpha 7$  ligands. *J Med Chem* 53:7192–7201
  16. Christopher JA, Brown J, Doré AS, Errey JC, Koglin M, Marshall FH, Myszka DG, Rich RL, Tate CG, Tehan B, Warne T, Congreve M (2013) Biophysical fragment screening of the  $\beta 1$ -adrenergic receptor: identification of high affinity arylpiperazine leads using structure-based drug design. *J Med Chem* 56:3446–3455
  17. Hesterkamp T, Barker J, Davenport A, Whittaker M (2007) Fragment based drug discovery using fluorescence correlation: spectroscopy techniques: challenges and solutions. *Curr Top Med Chem* 7:1582–1591
  18. Barker JJ, Barker O, Boggio R, Chauhan V, Cheng RK, Corden V, Courtney SM, Edwards N, Falque VM, Fusar F, Gardiner M, Hamelin EM, Hesterkamp T, Ichihara O, Jones RS, Mather O, Mercurio C, Minucci S, Montalbetti CA, Müller A, Patel D, Phillips BG, Varasi M, Whittaker M, Winkler D, Yarnold CJ (2009) Fragment-based identification of Hsp90 inhibitors. *ChemMedChem* 4:963–966
  19. Ganesh A (1998) Strategies for the dynamic integration of combinatorial synthesis and screening. *Angew Chem Int Ed* 37:2828–2831
  20. Lehn J-M, Eliseev AV (2001) Dynamic combinatorial chemistry. *Science* 291:2331–2332
  21. Ramström O, Lehn J-M (2002) Drug discovery by dynamic combinatorial libraries. *Nat Rev Drug Discov* 1:26–36
  22. Erlanson DA, Lam JW, Wiesmann C, Luong TN, Simmons RL, DeLano WL, Choong IC, Burdett MT, Flanagan WM, Lee D, Gordon EM, O'Brien T (2003) In situ assembly of enzyme inhibitors using extended tethering. *Nat Biotechnol* 21:308–314
  23. Choong IC, Lew W, Lee D, Pham P, Burdett MT, Lam JW, Wiesmann C, Luong TN, Fahr B, DeLano WL, McDowell RS, Allen DA, Erlanson DA, Gordon EM, O'Brien T (2002) Identification of potent and selective small-molecule inhibitors of caspase-3 through the use of extended tethering and structure-based drug design. *J Med Chem* 45:5005–5022
  24. Lewis WG, Green LG, Grynszpan F, Radić Z, Carlier PR, Taylor P, Finn MG, Sharpless KB (2002) Click chemistry in situ: acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks. *Angew Chem Int Ed* 41:1053–1057
  25. Krasinski A, Radić Z, Manetsch R, Raushel J, Taylor P, Sharpless KB, Kolb HC (2005) In situ selection of lead compounds by click chemistry: target-guided optimization of acetylcholinesterase inhibitors. *J Am Chem Soc* 127:6686–6692
  26. Whiting M, Muldoon J, Lin YC, Silverman SM, Lindstrom W, Olson AJ, Kolb HC, Finn MG, Sharpless KB, Elder JH, Fokin VV (2006) Inhibitors of HIV-1 protease by using in situ

- click chemistry. *Angew Chem Int Ed* 45: 1435–1439
27. Babaoglu K, Shoichet BK (2006) Deconstructing fragment-based inhibitor discovery. *Nat Chem Biol* 2:720–723
28. Barelier S, Pons J, Marcillat O, Lancelin J-M, Krimm I (2010) Fragment-based deconstruction of Bcl-xL inhibitors. *J Med Chem* 53:2577–2588
29. Van Molle I, Thomann A, Buckley DL, So EC, Lang S, Crews CM, Ciulli A (2012) Dissecting fragment-based lead discovery at the von Hippel-Lindau protein:hypoxia inducible factor 1 $\alpha$  protein-protein interface. *Chem Biol* 19:1300–1312
30. Krueger BA, Dietrich A, Baringhaus KH, Schneider G (2009) Scaffold-hopping potential of fragment-based *de novo* design: the chances and limits of variation. *Comb Chem High Throughput Screen* 12:383–396
31. Sun H, Tawa G, Wallqvist A (2012) Classification of scaffold-hopping approaches. *Drug Discov Today* 17:310–324
32. Ji H, Stanton BZ, Igarashi J, Li H, Martásek P, Roman LJ, Poulos TL, Silverman RB (2008) Minimal pharmacophoric elements and fragment hopping, an approach directed at molecular diversity and isozyme selectivity. Design of selective neuronal nitric oxide synthase inhibitors. *J Am Chem Soc* 130:3900–3914
33. Ji H, Li H, Martásek P, Roman LJ, Poulos TL, Silverman RB (2009) Discovery of highly potent and selective inhibitors of neuronal nitric oxide synthase by fragment hopping. *J Med Chem* 52:779–797
34. Lin F-Y, Tseng YJ (2011) Structure-based fragment hopping for lead optimization using predocked fragment database. *J Chem Inf Model* 51:1703–1715
35. Saluste G, Albaran MI, Alvarez RM, Rabal O, Ortega MA, Blanco C, Kurz G, Salgado A, Pevallo P, Bischoff JR, Pastor J, Oyarzabal J (2012) Fragment-hopping-based discovery of a novel chemical series of proto-oncogene PIM-1 kinase inhibitors. *PLoS One* 7:e45964
36. Yu B, Huang Z, Zhang M, Dillard DR, Ji H (2013) Rational design of small-molecule inhibitors for  $\beta$ -catenin/T-cell factor protein-protein interactions by bioisostere replacement. *ACS Chem Biol* 8:524–529
37. De Luca L, Ferro S, Morreale F, Christ F, Debysier Z, Chimirri A, Gitto R (2013) Fragment hopping approach directed at design of HIV IN-LEDGF/p75 interaction inhibitors. *J Enzyme Inhib Med Chem* 28:1002–1009
38. Goodford PJ (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 28:849–857
39. von Itzstein M, Wu W-Y, Kok GB, Pegg MS, Dyason JC, Jin B, Van Phan T, Smythe ML, White HF, Oliver SW, Colman PM, Varghese JN, Ryan DM, Woods JM, Bethell RC, Hotham VJ, Cameron JM, Penn CR (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 363:418–423
40. Halgren TA (2009) Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* 49:377–389
41. Miranker A, Karplus M (1991) Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* 11:29–34
42. Zoete V, Meuwly M, Karplus M (2005) Study of the insulin dimerization: binding free energy calculations and per-residue free energy decomposition. *Proteins* 61:79–93
43. Momany FA, Rone R (1992) Validation of the general purpose QUANTA® 3.2/CHARMM® force field. *J Comput Chem* 13:888–900
44. Kastenholz MA, Pastor M, Cruciani G, Haaksma EEJ, Fox T (2000) GRID/CPCA: a new computational tool to design selective ligands. *J Med Chem* 43:3033–3044
45. Ji H, Li H, Flinspach M, Poulos TL, Silverman RB (2003) Computer modeling of selective regions in the active site of nitric oxide synthases: implication for the design of isoform-selective inhibitors. *J Med Chem* 46:5700–5711
46. Brenke R, Kozakov D, Chuang G-Y, Beglov D, Hall D, Landon MR, Mattos C, Vajda S (2009) Fragment-based identification of druggable ‘hot spots’ of proteins using Fourier domain correlation techniques. *Bioinformatics* 25:621–627
47. Ngan CH, Bohnuud T, Mottarella SE, Beglov D, Villar EA, Hall DR, Kozakov D, Vajda S (2012) FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res* 40:W271–W275
48. Kortemme T, Kim DE, Baker D (2004) Computational alanine scanning of protein-protein interfaces. *Sci STKE* 2004(219):2
49. Koes D, Khouri K, Huang Y, Wang W, Bista M, Popowicz GM, Wolf S, Holak TA, Dömling A, Camacho CJ (2012) Enabling large-scale design, synthesis and validation of small molecule protein-protein antagonists. *PLoS One* 7:e32839
50. Koes DR, Camacho CJ (2012) Small-molecule inhibitor starting points learned from protein-protein interaction inhibitor structure. *Bioinformatics* 28:784–791
51. Koes DR, Camacho CJ (2012) PocketQuery: protein-protein interaction inhibitor starting

- points from protein-protein interaction structure. *Nucleic Acids Res* 40:W387–W392
52. Meireles LMC, Dömling AS, Camacho CJ (2010) ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery. *Nucleic Acids Res* 38:W407–W411
  53. Böhm H-J (1992) The computer program LUDI: a new method for the *de novo* design of enzyme inhibitors. *J Comput Aided Mol Des* 6:61–78
  54. Böhm H-J, Banner DW, Weber L (1999) Combinatorial docking and combinatorial chemistry: design of potent non-peptide thrombin inhibitors. *J Comput Aided Mol Des* 13:51–56
  55. Loving K, Salam NK, Sherman W (2009) Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *J Comput Aided Mol Des* 23:541–554
  56. Sándor M, Kiss R, Keseru GM (2010) Virtual fragment docking by Glide: a validation study on 190 protein-fragment complexes. *J Chem Inf Model* 50:1165–1172
  57. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791
  58. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. *Tetrahedron* 36:3219–3228
  59. Mooij WTM, Verdonk ML (2005) General and targeted statistical potentials for protein-ligand interactions. *Proteins* 61:272–287
  60. Korb O, Stützle T, Exner TE (2009) Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model* 49:84–96
  61. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46:3–26
  62. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 45:2615–2623
  63. Norinder U, Haeberlein M (2002) Computational approaches to the prediction of the blood-brain distribution. *Adv Drug Deliv Rev* 54:291–313
  64. Clark DE (2003) *In silico* prediction of blood-brain barrier permeation. *Drug Discov Today* 8:927–933
  65. Morelli X, Bourgeas R, Roche P (2011) Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr Opin Chem Biol* 15:475–481
  66. Cruciani G, Carosati E, De Boeck B, Ethirajulu K, Mackie C, Howe T, Vianello R (2005) MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J Med Chem* 48:6970–6979
  67. Wood WJL, Patterson AW, Tsuruoka H, Jain RK, Ellman JA (2005) Substrate activity screening: a fragment-based method for the rapid identification of nonpeptidic protease inhibitors. *J Am Chem Soc* 127: 15521–15527
  68. Patterson AW, Wood WJL, Ellman JA (2007) Substrate activity screening (SAS): a general procedure for the preparation and screening of a fragment-based non-peptidic protease substrate library for inhibitor discovery. *Nat Protoc* 2:424–433
  69. Greene J, Kahn S, Savoj H, Sprague P, Teig S (1994) Chemical function queries for 3D database search. *J Chem Inf Comput Sci* 34: 1297–1308
  70. Barnum D, Greene J, Smellie A, Sprague P (1996) Identification of common functional configurations among molecules. *J Chem Inf Comput Sci* 36:563–571
  71. Krovat EM, Langer T (2003) Non-peptide angiotensin II receptor antagonists: chemical feature based pharmacophore identification. *J Med Chem* 46:716–726
  72. Cottrell SJ, Gillet VJ, Taylor R, Wilton DJ (2004) Generation of multiple pharmacophore hypotheses using multiobjective optimisation techniques. *J Comput Aided Mol Des* 18: 665–682
  73. Richmond NJ, Abrams CA, Wolohan PRN, Abrahamian E, Willett P, Clark RD (2006) GALAHAD: I. Pharmacophore identification by hypermolecular alignment of ligands in 3D. *J Comput Aided Mol Des* 20:567–587
  74. Dixon SL, Smolydrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: I. Methodology and preliminary results. *J Comput Aided Mol Des* 20:647–671
  75. Sastry GM, Dixon SL, Sherman W (2011) Rapid shape-based ligand alignment and virtual screening method based on atom/feature-pair similarities and volume overlap scoring. *J Chem Inf Model* 51:2455–2466
  76. Wolber G, Langer T (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model* 45:160–169
  77. Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). I. Effect of shape on binding of

- steroids to carrier proteins. *J Am Chem Soc* 110:5959–5967
78. Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* 37:4130–4146
79. Cruciani G, Watson KA (1994) Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *J Med Chem* 37:2589–2601
80. Catana C (2009) Simple idea to generate fragment and pharmacophore descriptors and their implications in chemical informatics. *J Chem Inf Model* 49:543–548
81. Cecchini M, Kolb P, Majeux N, Caflisch A (2004) Automated docking of highly flexible ligands by genetic algorithms: a critical assessment. *J Comput Chem* 25:412–422
82. Huang D, Lüthi U, Kolb P, Cecchini M, Barberis A, Caflisch A (2006) *In silico* discovery of  $\beta$ -secretase inhibitors. *J Am Chem Soc* 128:5436–5443
83. Coleman RG, Carchia M, Sterling T, Irwin JJ, Shoichet BK (2013) Ligand pose and orientational sampling in molecular docking. *PLoS One* 8:e75992
84. Verdonk ML, Giangreco I, Hall RJ, Korb O, Mortenson PN, Murray CW (2011) Docking performance of fragments and druglike compounds. *J Med Chem* 54:5422–5431
85. Budin N, Majeux N, Caflisch A (2001) Fragment-based flexible ligand docking by evolutionary optimization. *Biol Chem* 382: 1365–1372

# Chapter 7

## Site Identification by Ligand Competitive Saturation (SILCS) Simulations for Fragment-Based Drug Design

**Christina E. Faller, E. Prabhu Raman, Alexander D. MacKerell, Jr., and Olgun Guvench**

### Abstract

Fragment-based drug design (FBDD) involves screening low molecular weight molecules (“fragments”) that correspond to functional groups found in larger drug-like molecules to determine their binding to target proteins or nucleic acids. Based on the principle of thermodynamic additivity, two fragments that bind nonoverlapping nearby sites on the target can be combined to yield a new molecule whose binding free energy is the sum of those of the fragments. Experimental FBDD approaches, like NMR and X-ray crystallography, have proven very useful but can be expensive in terms of time, materials, and labor. Accordingly, a variety of computational FBDD approaches have been developed that provide different levels of detail and accuracy.

The Site Identification by Ligand Competitive Saturation (SILCS) method of computational FBDD uses all-atom explicit-solvent molecular dynamics (MD) simulations to identify fragment binding. The target is “soaked” in an aqueous solution with multiple fragments having different identities. The resulting computational competition assay reveals what small molecule types are most likely to bind which regions of the target. From SILCS simulations, 3D probability maps of fragment binding called “FragMaps” can be produced. Based on the probabilities relative to bulk, SILCS FragMaps can be used to determine “Grid Free Energies (GFEs),” which provide per-atom contributions to fragment binding affinities. For essentially no additional computational overhead relative to the production of the FragMaps, GFEs can be used to compute Ligand Grid Free Energies (LGFEs) for arbitrarily complex molecules, and these LGFEs can be used to rank-order the molecules in accordance with binding affinities.

**Key words** Fragment-based drug design (FBDD), Molecular dynamics (MD), Site identification by ligand competitive saturation (SILCS), Binding free energy, FragMap, Grid free energy (GFE), Ligand grid free energy (LGFE)

---

### 1 General FBDD Methods and the SILCS Approach

Fragment-based drug design (FBDD) seeks to identify low molecular weight molecules (“fragments”) that bind to target proteins or nucleic acids of interest. The identities of these fragments are chosen based on their similarity to functional groups commonly occurring in drug-like molecules. After determining which of these

fragments bind, as well as their binding poses, fragments binding to adjacent sites on the target can be linked to create a molecule with a higher binding affinity [1]. This approach derives from the principle of thermodynamic additivity, which states that if two components are independent in their contributions to the change in free energy, then the sum of their respective contributions gives the total change in free energy, i.e.,  $\Delta G_{\text{total}} = \Delta G_{\text{fragment1}} + \Delta G_{\text{fragment2}}$  [2]. A variety of experimental techniques, including X-ray crystallography, NMR, surface plasmon resonance, isothermal titration calorimetry, and mass spectrometry, have proven to be very useful in determining binding affinities and binding poses of fragments to target proteins [3–6]. However, FBDD can be an expensive endeavor using experimental approaches, as they are associated with high costs in materials and time, especially for high-throughput screening, as well as in labor.

Computational approaches to FBDD aim to minimize the various costs associated with experimental approaches. Many *in silico* methods utilize simplified representations of the target and of the solvent in order to reduce the computational burden by reducing the number of degrees of freedom in the system. Examples of common simplifications include a rigid target model and representing the solvent as a continuum [7–11]. The rigid target model approach is often referred to as docking and has difficulty identifying ligands that require even minor changes in target conformation for binding [12–15]. More recent work has sought to improve sampling in this regard by using several different rigid target conformations for docking calculations [16]. Because of its computational speed, FBDD docking can allow high-throughput screening of large libraries of fragments that approach the theoretical limit of fragment diversity, which is  $10^7$  unique fragments [17]. FBDD docking, in addition to having the capacity to test all possible fragments, benefits from the fact that fragments have few internal degrees of freedom, which greatly simplifies the conformational search problem in docking [18–20]. However, development of sufficiently accurate scoring functions for ranking different docked molecules continues to be a challenge [21–24].

The opposite end of the spectrum from rigid target docking is the application of all-atom explicit-solvent molecular dynamics (MD) simulations, in which the solvent is explicitly modeled in atomic detail, and the ligands and target protein or nucleic acid are all fully flexible. In these MD simulations binding free energies can be determined and, in conjunction with the optimized empirical force fields presently available for biomolecules and small molecules [25–33], near-quantitative binding free energy agreement can be reached relative to wet-lab experiments [34–43]. Unfortunately, while this level of detail provides accuracy, computational efficiency is lost due to the need to sample ligand, target, and solvent degrees of freedom sufficiently to obtain converged

results. Therefore, using this type of thorough MD simulation to do high-throughput analysis of fragment binding is simply not possible for the foreseeable future.

While it is computationally inefficient to do MD simulations on individual small molecules from a large set, a possible advantageous approach is to employ a competitive method that simultaneously screens a simplified set of fragment molecules that represent various functional groups. The SILCS (Site Identification by Ligand Competitive Saturation) method [44] does exactly this by using atomic level of detail MD simulations of a target in an aqueous solution containing selected fragment molecules so as to determine regions of high probability binding for different fragment types.

---

## 2 SILCS Methodological Details

SILCS [44] uses nanosecond-length all-atom explicit-solvent MD simulations of the target in an aqueous solution containing a variety of fragments. Explicitly modeling water molecules allows for atomic-level solvation effects to be included. Multiple simulations are run for each system, the trajectories are combined, and 3D probability maps of each fragment type around the target are calculated. The 3D probability maps are then normalized relative to fragment probabilities in bulk solution, thereby incorporating fragment desolvation free energies into the final maps, which are referred to as “FragMaps.” As explicit water is included in the MD simulations, the free energy penalty for desolvation of the target to allow fragments binding is taken into account in the final FragMaps in addition to all other components of binding free energy including target–ligand interactions, target deformation energy, and entropic contributions.

### 2.1 Fragment Selection

Selecting fragments to include in the aqueous solution is an important step in the SILCS methodology: the fragments should be small enough to allow adequate concentrations to facilitate conformational sampling, and should minimally represent hydrogen bond donors, hydrogen bond acceptors, aliphatic groups, and aromatic groups. In the original conception of SILCS, the water portion of the solution contributed both hydrogen bond donors and hydrogen bond acceptors. The other fragments, therefore, needed to include aliphatic and aromatic moieties, meaning two more fragments were required to provide these functional moieties to complement those provided by water. Low molecular weight fragments are particularly desirable due to their high diffusion rates, which lead to improved convergence of simulations. By these standards, the original SILCS fragments were water, benzene, and propane. However, while water is a convenient choice, there are potentially many other hydrogen bond donor and/or hydrogen

bond acceptor-containing fragment choices that are more representative of moieties in drug-like molecules. As such, the original SILCS fragment set is now referred to as “Tier 1,” and a new “Tier 2” fragment set consisting of propane, benzene, methanol, formamide, acetaldehyde, methylammonium, and acetate has been validated [45]. Notable in Tier 2 SILCS is the use of both neutral and charged donors and acceptors, allowing for regions of the target that bind these to be differentiated.

Low molecular weight fragments have an additional benefit, which stems from the competitive nature of the SILCS in silico assay: there is an upper limit to the ligand binding affinity per heavy atom [46], commonly referred to as “ligand efficiency” [47], and this limit is 0.4–0.5 kcal/mol per heavy atom [17]. As a consequence, smaller fragments (fewer heavy atoms) translate to weaker binding and higher turnover of fragments on target binding sites, which improves sampling. Characterization of such weakly binding fragments through NMR and X-ray crystallography experiments can be challenging, which limits the number of fragment types that are recognized in experimental FBDD efforts. In contrast, SILCS does not have this limitation. Of course, fragments other than those mentioned above can be used to further broaden the range of chemical space represented by FragMaps, but again it is emphasized that larger fragments may slow convergence because of slower diffusion and greater binding affinity.

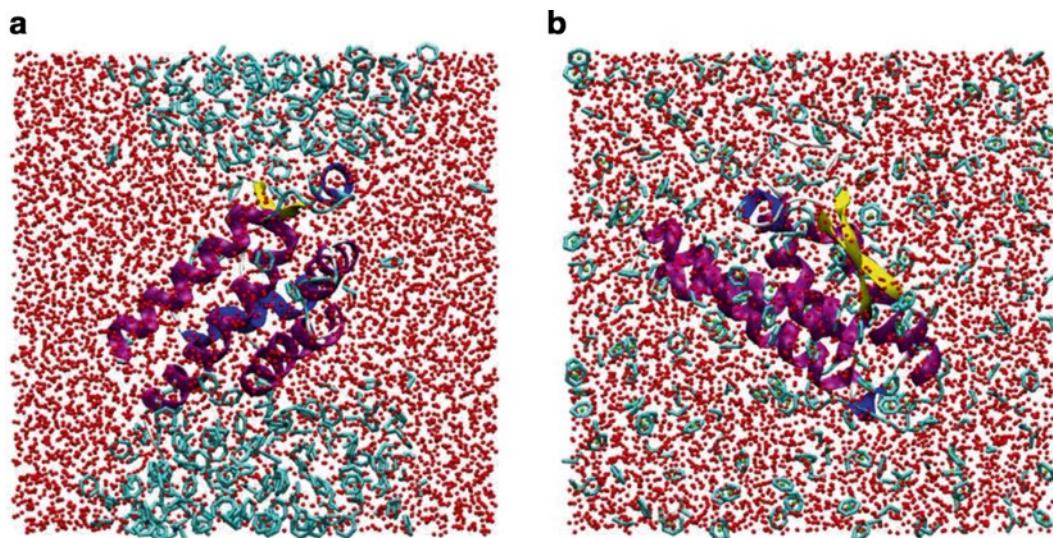
## 2.2 Preventing Fragment Aggregation

The fact that SILCS uses fragment concentrations approaching 1 M in an aqueous solution brings about the problem of aggregation. The aggregation of hydrophobic molecules occurs because they prefer not to be solvated, but to associate with other hydrophobic molecules. The resulting phase separation has the serious consequence that the effective concentration of the hydrophobic molecules is substantially reduced. In Tier 2 SILCS, the presence of ions can lead to ion-pair formation in solution, again reducing the effective fragment concentration. In both instances, the chemical potential of the fragments in bulk solution is reduced, thereby reducing their sampling of the target surface.

SILCS overcomes this barrier to sampling by leveraging the fact that it is a computational method: in SILCS, a repulsive potential between fragments is used to prevent fragment association [44]. This repulsive potential—unique to SILCS—only alters fragment–fragment interactions in the system, thereby maintaining an “ideal” solution of fragments in water while leaving all other interactions in the system unperturbed (Fig. 1).

## 2.3 Balancing Target Flexibility and Denaturation

The value of including target flexibility can be visualized by comparing atomic resolution structures of apo- and ligand-bound proteins; in many cases, binding of a ligand is coupled to a conformational change, including in therapeutically relevant targets such as kinases



**Fig. 1** Fragment aggregation after 20 ns of SILCS MD simulation. **(a)** No inter-fragment repulsive potential. **(b)** With SILCS inter-fragment repulsive potential. The protein target is displayed as ribbons, water oxygen atoms are in *red*, and benzene and propane carbon atoms are in *blue*

and proteases [48]. Common to SILCS-like approaches [49–53] is the risk of fragment-induced target denaturation, especially in cases of inherently less-stable target proteins, such as those with no disulfide bonds. A range of options has been considered with regard to treatment of target flexibility in SILCS to optimally balance inclusion of target flexibility while minimizing the risk of fragment-induced target denaturation: a fully flexible protein (no positional restraints), weak C $\alpha$  positional restraints, and weak positional restraints on non-hydrogen atoms near the protein core [54]. The weak positional restraints allow for relatively large motions of the restrained atoms while limiting the motions enough to avoid denaturation. What is clear is that a lack of restraints (i.e., “full flexibility”) can allow target denaturation regardless of whether the fragments are hydrophobic or hydrophilic.

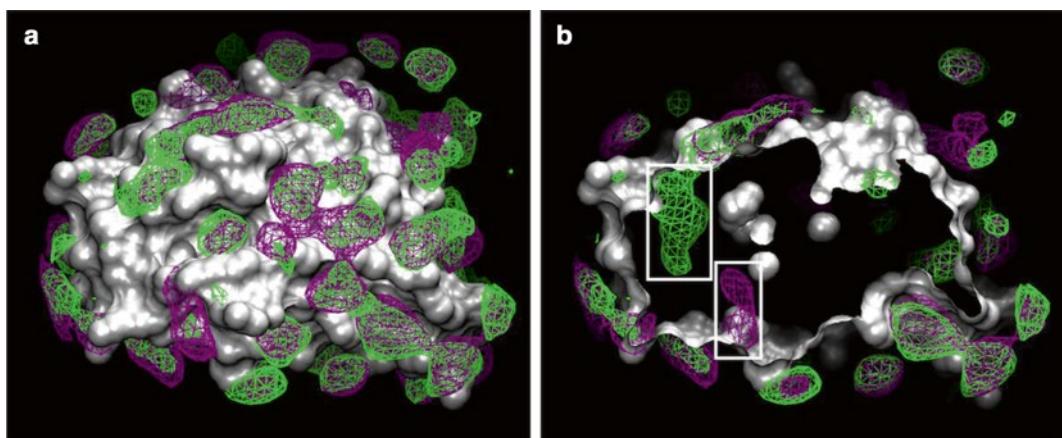
Should full target flexibility be required, a protocol has been developed to identify denaturing SILCS trajectories for exclusion from subsequent analysis. It employs a combined metric consisting of the root-mean square deviation (RMSD) and the radius of gyration ( $R_{\text{gyr}}$ ). The average RMSD (with the starting structure as reference coordinates) and the average  $R_{\text{gyr}}$  are computed for each SILCS trajectory. Likewise, they are computed for non-SILCS standard MD control trajectories of the target in the absence of fragments. Each SILCS trajectory average RMSD, average  $R_{\text{gyr}}$  pair is then compared to the cluster of these values from the control trajectories, and the SILCS trajectory is excluded from further

analysis if its values lie outside the cluster for the control simulation [54]. While large RMSD values alone may be good indicators of denaturation, intermediate values can require visualization of snapshots from the trajectory to confirm the RMSD results. In cases of structural rearrangement, like loop rearrangements or sliding of helices relative to one another, intermediate RMSD can be misleading, as the increase in RMSD is not due to denaturation, but due to functionally-relevant conformational change. The use of  $R_{\text{gyr}}$  has a long history as a reaction coordinate in computational studies of protein folding [55–57] and is a metric of the overall spatial extent of the protein that increases as a protein unfolds. In the context of SILCS,  $R_{\text{gyr}}$  is especially capable of identifying situations where fragments tunnel into and disrupt the protein hydrophobic core, which may lead to ambiguous intermediate changes in RMSD.

## 2.4 FragMaps

FragMaps are 3D probability distributions of the fragment atom types in the context of the target. They serve to identify which functionalities (e.g., hydrogen bond donors, hydrogen bond acceptors, aromatic groups, aliphatic groups) associate most strongly with different areas of the target. FragMaps can be conveniently visualized as isocontour surfaces in the context of the target using freely available molecular graphics software like VMD [58] (Fig. 2) or PMV/ADT [59, 60], or any of the widely used commercial molecular visualization software packages, since FragMaps can be stored in the same formats as those used for electron densities [61] or docking grids [60].

In the initial SILCS implementation using Tier 1 fragments, to be included in a FragMap a fragment atom must meet a distance



**Fig. 2** Tier 1 FragMaps. (a) Target protein molecular surface in white, propane FragMap in *green* mesh, and benzene FragMap in *purple* mesh. (b) Same as (a) but with clipping and depth-cueing to expose additional FragMap density (*white boxes*) beneath the molecular surface of the apo-target crystal structure; this density corresponds to experimentally known ligand-binding sites [54]

criterion relative to the target protein: for example, for inclusion in the hydrogen bond acceptor FragMap, water molecule oxygen atoms must be within 2.5 Å of the protein [44]. This criterion was particularly relevant in order to distinguish whether the water was acting in a hydrogen bond donor or acceptor capacity, as a target-bound water molecule may be serving in either role or simultaneously in both roles. With the move to the more varied and specific Tier 2 fragments, this is less of an issue, and inclusion of all fragments during FragMap generation, regardless of distance from the target, can be useful to capture longer-range interactions such as water-mediated interactions of polar molecules with the target. In either of these two approaches, fragment atom locations are binned to create a 3D histogram (FragMap) having 1 Å×1 Å×1 Å voxels.

## 2.5 Determining Convergence

A practical means to evaluate SILCS simulation convergence is to run ten independent simulations, and create two sets of FragMaps by averaging over two sets of five independent FragMaps. Data in the second set are combined and subtracted from the combined data in the first set to generate a difference map. If the simulations are converged, differences should be due to random error and therefore have a tight distribution centered around zero [44]. Alternatively, the overlap coefficient of the two maps can be calculated to gauge the extent of convergence [62].

## 2.6 Grid Free Energies (GFEs)

To obtain quantitative free-energy information, FragMaps are normalized relative to fragment occupancies in bulk solvent and converted to “Grid Free Energies (GFEs)” via inverse-Boltzmann weighting of the normalized FragMap occupancies [63]. In order to normalize results, simulations with conditions similar to those of the target+fragments+water system are run with only the SILCS solution (i.e., water+fragments). As with the target-containing simulations, the solution-only simulations are run in the isothermal-isobaric (*NPT*) ensemble to allow for system size relaxation to account for the volume occupied by fragments in the solution. After relaxation, the bulk occupancy for a particular fragment type is computed by simply dividing the number of atoms of that fragment type by the average volume of the system computed from the *NPT* simulations. The GFE for a fragment atom type *f* in a particular voxel centered at *x*, *y*, *z* is then,

$$\text{GFE}_{x,y,z}^f = \min \left\{ -RT \log_e \frac{\text{occupancy}_{x,y,z}^f}{\text{bulk occupancy}}, \text{GFE}_{\max} \right\} \quad (1)$$

where  $\text{GFE}_{\max}$  can be set as a maximum unfavorable value. A  $\text{GFE}_{\max}$  of 0 has been used previously [63], which removes any unfavorable contributions arising from voxels having an occupancy lower than bulk.

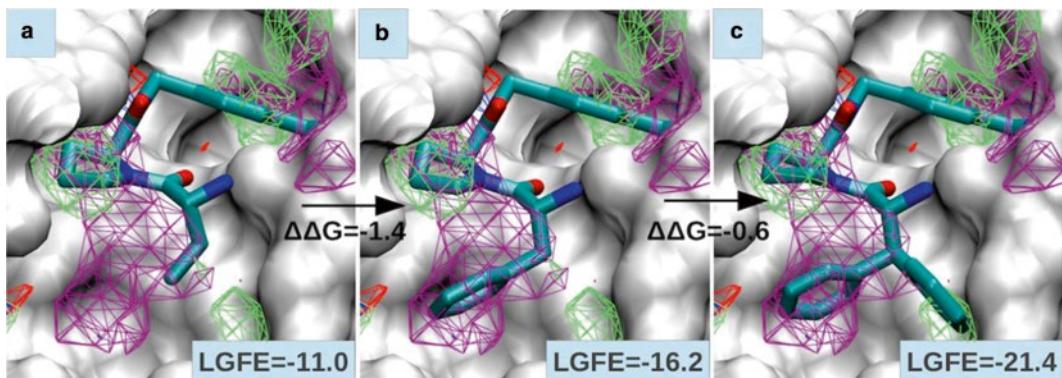
## 2.7 Ligand Grid Free Energy Scores (LGFEs)

Ligand Grid Free Energy scores (LGFEs) can provide an estimate of binding free energy of a particular target–ligand conformation for an arbitrarily complex ligand molecule. In order to calculate LGFE scores, ligand molecule atoms are classified into FragMap types based on their chemical similarity to the various fragment atoms used to compute the FragMaps. To this end, an assignment convention has been developed that translates force-field atom types into the FragMap classes. Additionally, certain ligand atoms may be excluded from the LGFE calculation. For example, aromatic hydrogen atoms are implicitly accounted for in the FragMap for the parent benzene carbon atom. The LGFE for a ligand molecule is computed as a sum of the GFEs of its classified atoms:

$$\text{LGFE} = \sum_{\text{FragMaps } f} \sum_{\text{atoms } i_f} GFE_{x,y,z}^f(i_f) \quad (2)$$

where the outer summation is over the FragMap types denoted by  $f$  and the inner summation is over the atoms denoted by  $i_f$  that are classified into each FragMap type. In addition to single conformations, LGFE scores can be computed for an ensemble of conformations and thermodynamically averaged. Such ensembles of conformations may be obtained, for example, from relatively short (e.g., 1–2 ns) Langevin simulations of the target–ligand complex in the gas phase or with a continuum solvent model. When this is performed it is suggested that the simulations be repeated multiple times with different target conformations.

Figure 3 demonstrates the utility of LGFE scores in structure based drug design. The crystallographic conformations of three ligands that bind to the protein  $\alpha$ -thrombin with progressively increasing affinities are shown, along with LGFE scores and experimentally



**Fig. 3** Crystallographic complexes of  $\alpha$ -thrombin with three ligands of progressively higher affinity, along with benzene and propane FragMaps. The benzene and propane FragMaps are in purple and green, respectively. The ligand grid free energy (LGFE) for each ligand is displayed on the right-bottom side of each panel and the experimentally measured binding affinity difference is at the interface of each pair of panels. Protein–ligand structures are from PDB IDs (a) 2ZGX, (b) 2ZDA, and (c) 2Z03

determined binding free energies. The overlap of the optimized ligands with FragMaps reflected in the increasingly favorable LGFE scores captures the experimental trend of increasing binding affinities [63, 64].

### 3 SILCS Workflow

#### 3.1 System Construction

1. Determine size of system: a cube with edge lengths  $x$  is typical, where  $x$  is 16 Å longer than the longest dimension of the target.
2. Generate a box of water molecules of edge length  $x$  at the experimental density.
3. Choose a fragment palette, e.g.: Tier 1 SILCS=propane and benzene; Tier 2 SILCS=propane, benzene, methanol, formamide, acetaldehyde, methylammonium, and acetate.
4. Compute fragment placement grid overlapping with water box, such that a ~1 M for Tier 1 or ~0.25 M for Tier 2 solution in each of the fragments will result.
5. At each placement grid point, randomly select and place a fragment from the palette to generate a fragment solution box.
6. Center the target in fragment solution box from above.
7. Delete water molecules and fragments overlapping with target.

#### 3.2 System Simulation

1. Nonbonded conditions: 8 Å real-space cutoff; Particle-mesh Ewald for long-range electrostatics; Switching function between 5 and 8 Å for Lennard-Jones; Virtual particles are added to center of each fragment to serve as interaction sites for inter-fragment repulsion (repulsive potential between the virtual particle pairs is modeled using Lennard-Jones potential combined with above switching function, where Lennard-Jones parameters are  $e = -0.01$  kcal/mol,  $R_{\min} = 12.0$  Å).
2. Positional restraints for protein target sidechain flexibility: Harmonic restraints on C $\alpha$  atom positions of the form  $k(\Delta r)^2$ , where  $\Delta r$  is the displacement in Å from the crystallographic position and  $k = 0.1$  kcal/mol/Å $^2$ . In addition to full sidechain flexibility, these restraints are sufficiently weak to allow modest backbone flexibility.
3. *Or* positional restraints for loop flexibility: Harmonic restraints on non-hydrogen atoms within  $x$  Å of the target center of mass, where  $x$  is sufficiently small (e.g., half of the radius of gyration of the target) so as not to include residues near the surface of the target. Restraint functional form and force constants are same as for “sidechain flexibility” above.

4. *Or* full flexibility: No positional restraints. In this case, run control simulations of just the target without fragments for use in post-run determination of denaturing trajectories using average RMSD, average  $R_{\text{gyr}}$  pair metric.
5. Simulate with isothermal-isobaric (*NPT*) molecular dynamics (MD).
6. Typical simulation length is 50 ns.
7. It is recommended that ten independent simulations be run: Same solution box with ten different random seeds to initiate MD, or, preferably, ten different solution boxes.

### **3.3 FragMap Generation**

1. For a given fragment atom type  $f$ , bin all atomic positions from all SILCS trajectory snapshots to create a 3D histogram spanning the size of the system and having  $1 \text{ \AA} \times 1 \text{ \AA} \times 1 \text{ \AA}$  voxels.
2. If, as recommended, multiple independent simulations were run, generate two FragMaps for each fragment atom type  $f$ , using half the simulations for each FragMap, and compute a difference map or overlap coefficient to estimate convergence.
3. n.b.: If “full flexibility” was used, snapshots from denaturing trajectories must be excluded and each snapshot needs to be aligned to a single reference orientation of the target prior to binning of fragment atom positions.

### **3.4 GFE Generation**

1. See Eq. 1.

### **3.5 LGFE Computation**

1. Ligand conformation generation: Can be a docking pose, multiple poses from an MD simulation of the target–ligand complex, etc.
2. Ligand atom classification: Each ligand atom is mapped to a fragment atom type  $f$  based on chemical similarity.
3. LGFE is computed using Eq. 2.

## **Acknowledgements**

NIH AI080968, CA107331, and R15GM099022; NSF XSEDE TG-MCB120007; Samuel Waxman Cancer Research Foundation; The University of Maryland Computer-Aided Drug Design Center; and University of New England start-up funds.

*Conflict of interest:* O.G. and A.D.M. are founders of SilcsBio LLC and are currently Manager and Chief Scientific Officer of the same, respectively.

## References

1. Erlanson DA, McDowell RS, O'Brien T (2004) Fragment-based drug discovery. *J Med Chem* 47:3463–34682
2. Dill KA (1997) Additivity principles in biochemistry. *J Biol Chem* 272:701–704
3. Hennig M, Ruf A, Huber W (2012) Combining biophysical screening and X-ray crystallography for fragment-based drug discovery. *Top Curr Chem* 317:115–143
4. Erlanson DA (2012) Introduction to fragment-based drug discovery. *Top Curr Chem* 317:1–32
5. Orita M, Warizaya M, Amano Y, Ohno K, Niimi T (2009) Advances in fragment-based drug discovery platforms. *Expert Opin Drug Discov* 4:1125–1144
6. Sancinetto L, Massari S, Iraci N, Tabarrini O (2013) From small to powerful: the fragments universe and its “chem-appeal”. *Curr Med Chem* 20:1355–1381
7. Miranker A, Karplus M (1991) Functionality maps of binding-sites – a multiple copy simultaneous search method. *Proteins Struct Funct Genet* 11:29–34
8. Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A (1999) Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins Struct Funct Genet* 37:88–105
9. Landon MR, Lancia DR Jr, Yu J, Thiel SC, Vajda S (2007) Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J Med Chem* 50:1231–1240
10. Clark M, Guarnieri F, Shkurko I, Wiseman J (2006) Grand canonical Monte Carlo simulation of ligand-protein binding. *J Chem Inf Model* 46:231–242
11. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, Briggs JM, McCammon JA (2000) Developing a dynamic pharmacophore model for HIV-1 integrase. *J Med Chem* 43:2100–21014
12. Alonso H, Bliznyuk AA, Gready JE (2006) Combining docking and molecular dynamic simulations in drug design. *Med Res Rev* 26:531–568
13. Brooijmans N, Kuntz ID (2003) Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* 32:335–373
14. Sousa SF, Fernandes PA, Ramos MJ (2006) Protein-ligand docking: current status and future challenges. *Proteins Struct Funct Bioinf* 65:15–26
15. Zhong S, Chen X, Zhu X, Dziegielewska B, Bachman KE, Ellenberger T, Ballin JD, Wilson GM, Tomkinson AE, MacKerell AD (2008) Identification and validation of human DNA ligase inhibitors using computer-aided drug design. *J Med Chem* 51:4553–4562
16. Totrov M, Abagyan R (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol* 18:178–184
17. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent developments in fragment-based drug discovery. *J Med Chem* 51:3661–3680
18. Majeux N, Scarsi M, Caflisch A (2001) Efficient electrostatic solvation model for protein-fragment docking. *Proteins Struct Funct Genet* 42:256–268
19. Gozalbes R, Carbajo RJ, Pineda-Lucena A (2010) Contributions of computational chemistry and biophysical techniques to fragment-based drug discovery. *Curr Med Chem* 17:1769–1794
20. Rabal O, Urbano-Cuadrado M, Oyarzabal J (2011) Computational medicinal chemistry in fragment-based drug discovery: what, how and when. *Future Med Chem* 3:95–134
21. Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR (2008) Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br J Pharmacol* 153(Suppl 1):S7–S26
22. Guvench O, MacKerell AD Jr (2009) Computational evaluation of protein-small molecule binding. *Curr Opin Struct Biol* 19:56–61
23. Huang SY, Grinter SZ, Zou X (2010) Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys* 12:12899–12908
24. Huang SY, Zou X (2010) Advances and challenges in protein-ligand docking. *Int J Mol Sci* 11:3016–3034
25. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* 25:1656–1676
26. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197
27. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison

- of multiple amber force fields and development of improved protein backbone parameters. *Proteins Struct Funct Bioinf* 65:712–725
- 28. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174
  - 29. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118:11225–11236
  - 30. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105:6474–6487
  - 31. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD Jr (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 31:671–690
  - 32. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, MacKerell AD Jr (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput* 8:3257–3273
  - 33. Guvench O, MacKerell AD Jr (2008) Comparison of protein force fields for molecular dynamics simulations. *Methods Mol Biol* 443:63–88
  - 34. Boresch S, Tettinger F, Leitgeb M, Karplus M (2003) Absolute binding free energies: a quantitative approach for their calculation. *J Phys Chem B* 107:9535–9551
  - 35. Jayachandran G, Shirts MR, Park S, Pande VS (2006) Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *J Chem Phys* 125:084901
  - 36. Jiao D, Golubkov PA, Darden TA, Ren P (2008) Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc Natl Acad Sci U S A* 105:6290–6295
  - 37. Lee MS, Olson MA (2006) Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys J* 90: 864–877
  - 38. Lee MS, Olson MA (2008) Calculation of absolute ligand binding free energy to a ribosome-targeting protein as a function of solvent model. *J Phys Chem B* 112:13411–13417
  - 39. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA (2007) Predicting absolute ligand binding free energies to a simple model site. *J Mol Biol* 371: 1118–1134
  - 40. Shirts MR, Mobley DL, Chodera JD, Pande VS (2007) Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J Phys Chem B* 111: 13052–13063
  - 41. Wang J, Deng Y, Roux B (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys J* 91:2798–2814
  - 42. Woo HJ, Roux B (2005) Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc Natl Acad Sci U S A* 102:6825–6830
  - 43. Gilson MK, Zhou HX (2007) Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct* 36:21–42
  - 44. Guvench O, MacKerell AD Jr (2009) Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comput Biol* 5:e1000435
  - 45. Raman EP, Yu W, Lakkaraju SK, MacKerell AD Jr (2013) Inclusion of multiple fragment types in the site identification by ligand competitive saturation (SILCS) approach. *J Chem Inf Model* 53:3384–3398
  - 46. Kuntz ID, Chen K, Sharp KA, Kollman PA (1999) The maximal affinity of ligands. *Proc Natl Acad Sci U S A* 96:9997–10002
  - 47. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* 9:430–431
  - 48. Spyrosakis F, BidonChanal A, Barril X, Luque FJ (2011) Protein flexibility and ligand recognition: challenges for molecular modeling. *Curr Top Med Chem* 11:192–210
  - 49. Seco J, Luque FJ, Barril X (2009) Binding site detection and druggability index from first principles. *J Med Chem* 52:2363–2371
  - 50. Yang C-Y, Wang S (2011) Hydrophobic binding hot spots of Bcl-xL protein, protein interfaces by cosolvent molecular dynamics simulation. *ACS Med Chem Lett* 2:280–284
  - 51. Lexa KW, Carlson HA (2011) Full protein flexibility is essential for proper hot-spot mapping. *J Am Chem Soc* 133:200–202
  - 52. Tan YS, Sledz P, Lang S, Stubbs CJ, Spring DR, Abell C, Best RB (2012) Using ligand-mapping simulations to design a ligand selectively targeting a cryptic surface pocket of polo-like kinase 1. *Angew Chem Int Ed Engl* 51:10078–10081
  - 53. Bakan A, Nevins N, Lakdawala AS, Bahar I (2012) Druggability assessment of allosteric proteins by dynamics simulations in the

- presence of probe molecules. *J Chem Theory Comput* 8:2435–4247
- 54. Foster TJ, Mackerell AD Jr, Guvench O (2012) Balancing target flexibility and target denaturation in computational fragment-based inhibitor discovery. *J Comput Chem* 33:1880–1891
  - 55. Boczek EM, Brooks CL III (1995) First-principles calculation of the folding free energy of a three-helix bundle protein. *Science* 269:393–396
  - 56. Sheinerman FB, Brooks CL III (1998) Calculations on folding of segment B1 of streptococcal protein G. *J Mol Biol* 278:439–456
  - 57. Shea JE, Brooks CL III (2001) From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu Rev Phys Chem* 52:499–535
  - 58. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38
  - 59. Sanner MF (1999) Python: a programming language for software integration and development. *J Mol Graph Model* 17:57–61
  - 60. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791
  - 61. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235–242
  - 62. Bernard D, Coop A, MacKerell AD Jr (2007) Quantitative conformationally sampled pharmacophore for delta opioid ligands: reevaluation of hydrophobic moieties essential for biological activity. *J Med Chem* 50:1799–1809
  - 63. Raman EP, Yu W, Guvench O, MacKerell AD Jr (2011) Reproducing crystal binding modes of ligand functional groups using site-identification by ligand competitive saturation (SILCS) simulations. *J Chem Inf Model* 51:877–896
  - 64. Baum B, Muley L, Heine A, Smolinski M, Hangauer D, Klebe G (2009) Think twice: understanding the high potency of bis(phenyl) methane inhibitors of thrombin. *J Mol Biol* 391:552–564

# Chapter 8

## A Computational Fragment-Based De Novo Design Protocol Guided by Ligand Efficiency Indices (LEI)

Álvaro Cortés-Cabrera, Federico Gago, and Antonio Morreale

### Abstract

We present a new protocol aimed at the structure-based design of drug-like molecules using a fragment approach. It starts from a suitably placed and well-defined “base fragment” and then uses an incremental construction algorithm and a scoring function to grow the molecule into prioritized candidates. The selection of the most promising solutions for synthesis and validation is guided by the optimization of the calculated ligand efficiency indices known as binding efficiency index (BEI) and surface efficiency index (SEI), which allow the user to navigate proficiently in chemico-biological space. A test case for the protocol is exemplified here using published data for inhibitors of protein kinase B, aka AKT, a key enzyme in several signal transduction pathways. Our procedure was able to identify the main features responsible for the binding of inhibitors and guided the selection process towards molecules that included or resembled those shown as the most active in the original studies.

**Key words** Fragment-based drug design, Docking, Ligand efficiency indices, Scoring functions

---

### 1 Introduction

Fragment-based drug design (FBDD) is a mature and well-established approach for drug discovery and optimization [1]. However, several limitations still exist related to the equipment and the expensive materials that are needed for the implementation of the experimental protocols. For this reason chemoinformatics and computational tools can assist those discovery efforts, in a parallel or independent manner, by simplifying the fragment space that needs to be explored or by pointing out which are the best spots within this space [2].

Although the definition of molecular fragment varies across the literature and depends on its intended use, the most common one takes into account size and physicochemical properties. Thus, the Rule of Three [3] states that the most successful fragments have a molecular weight (MW) of less than 300 kDa, a  $c\log P$  equal or less than 3 and a number of hydrogen bond donor and

acceptor atoms of less than or equal to 3. This definition has been applied widely, but it should not be considered in absolute terms because some successful studies have employed fragments that do not fulfill one or even two of these recommendations [4]. Fragment databases are generally obtained either directly from chemical suppliers that provide diverse sets of building blocks or from filtering some commonly used large chemical libraries using the above mentioned rule of three. Another, and perhaps more interesting, alternative is to break down the molecules present in drug-like databases into smaller pieces following a rational fragmentation scheme.

The concept of ligand efficiency was introduced to account for the differences in ligand affinity or potency with respect to molecular size [5]. As a consequence it is now common practice to normalize the binding free energy of a ligand with respect to different properties such as MW, number of heavy atoms, and polar surface area (PSA). The resulting ligand efficiency indices (LEIs) [6] have demonstrated to be very useful in both experimental and computational FBDD campaigns [7] and to properly describe the chemico-biological space (CBS) that is being explored for fragment optimization [8]. Prospective and retrospective analyses [9, 10] have shown that a given optimization path in CBS can be successfully predicted and followed using a LEI framework and 2D planes.

In the next sections we describe a fully LEI-driven computational protocol that employs a succinct and diverse fragment library together with a growing scheme within the binding site to suggest new target-oriented compounds with drug-like properties starting from a previously identified scaffold. The protocol makes use of a tailor-made four-module toolbox to perform the following tasks:

1. Binding pocket analysis to characterize the structural and energetic properties that are needed in subsequent steps.
2. Placement of a suitable base fragment that displays good steric and electrostatic complementarities with the binding site at a particular location.
3. Attachment of new fragments to the base fragment using a growing algorithm.
4. Scoring and optimization of the resulting candidates using a scoring function that maximizes the square sum of the binding efficiency index (BEI) and the surface efficiency index (SEI).

A Graphical User Interface (GUI), implemented as a plugin for the molecular visualization and editing program PyMOL [11], and two simple scripts facilitate the use of this toolchest by nonexperts (*see Note 1*).

The whole procedure is exemplified here for the target protein kinase B (PKB), a serine/threonine kinase that regulates many signaling pathways involved in cell growth and differentiation.

---

## 2 Materials

The co-crystal structures of a PKA-PKB chimeric protein in complex with every intermediate obtained during the optimization process in an experimentally validated FBDD campaign that also provided inhibitory activity values ( $IC_{50}$ ) for all compounds [12].

### 2.1 Three-Dimensional Coordinates for Protein and Fragments

1. Protein Data Bank (PDB) entry 2UW3.
2. Four databases containing fragments (stored as individual files) whose functionalities are intended to match the four different types of hotspots detected by the binding pocket analysis tool, i.e., hydrogen bond donors, hydrogen bond acceptors, mixed hydrogen bond donor/acceptor groups, and hydrophobic moieties (see Note 2).

### 2.2 Software

1. The binding pocket analysis tool cGRILL, a stand-alone C program (see Note 3) that is formally equivalent to Goodford's program GRID [13].
2. Two scripts for (a) library linking (*linker.py*) and (b) scoring new molecules (*scorer.py*).
3. Scoring functions MM-ISMSA [14], of general use and our default procedure (see Note 4), ChemScore [15], better suited for non-charged ligands and hydrophobic pockets (see Note 5), and HYDE [16], which works best for accurately placed ligands and can be applied to any protein–ligand/fragment structure (see Note 6).
4. A LEI-driven algorithm that calculates BEI and SEI for each candidate molecule and then plots the BEI vs. SEI efficiency plane [17] to guide the growing scheme (see Note 7).

---

## 3 Methods

### 3.1 Target Setup

1. Extract the protein coordinates (ATOM records) from PDB entry 2UW3 to create target file rec.pdb.
2. Add missing hydrogen atoms to all residues possibly taking into account the ionization state of titratable amino acids at the reference pH value (see Note 8).
3. For all ATOM records replace the occupancy and *B*-factor values (last two columns) in file rec.pdb with atomic radii (*R*) and atom point charges (*Q*), respectively, and save the resulting new file as rec.prq (PRQ or “swapped” PQR format) (see Note 8).

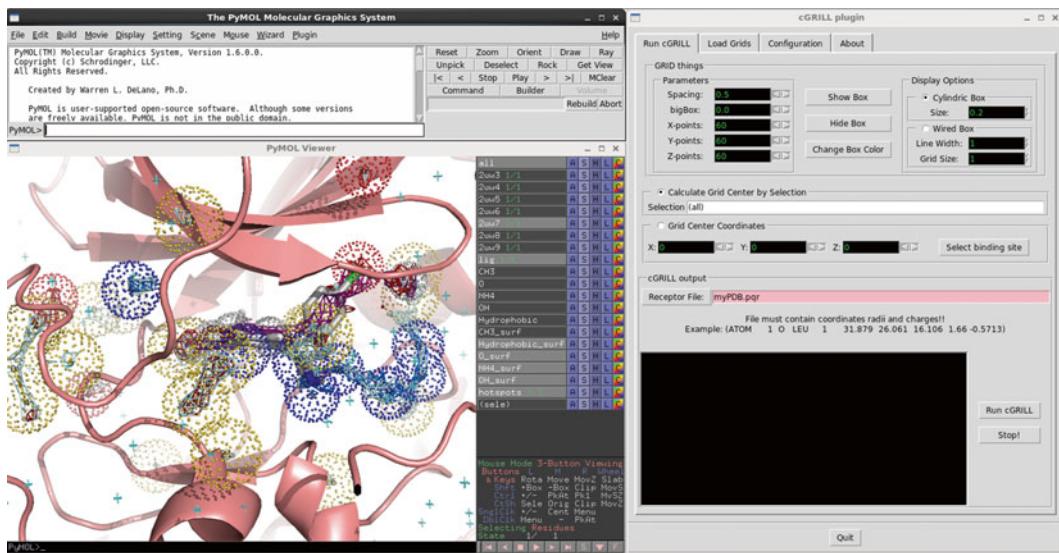
### 3.2 Base Fragment Setup

1. Extract residue GVG (3-methyl-4-phenyl-1*H*-pyrazole), with HETATM records, from PDB entry 2UW3 to create base fragment file lig.pdb.

2. If no experimentally determined hit exists for the target, perform a virtual screening campaign to identify a suitable starting scaffold as the “base fragment” and determine its optimal positioning within the binding pocket [18] (*see Note 9*).
3. Add any missing hydrogen atoms to lig.pdb (e.g., using OpenBabel or PyMOL, *see Note 10*), and write out the resulting file in PDB format as ligH.pdb.

### 3.3 Binding Pocket Characterization

1. In PyMOL, open the rec.pdb and ligH.pdb files to work on and the original 2UW3.pdb file for reference.
2. Display the cGRILL window by clicking on the corresponding option from the Plugin main menu.
3. Choose the *Configuration tab* to type in the path to the cGRILL executable and to specify the name of the working directory where results will be stored.
4. To define the grid center, select object ligH (the chosen base fragment) in the PyMOL Viewer window and type *sele* inside the *Selection* box under the *Run cGRILL* tab in the cGRILL plugin window. Press *Enter*.
5. Once the grid center coordinates and the cubic box chosen for the analysis are displayed, adjust the default dimensions ( $40 \times 40 \times 40$  points) and spacing (0.5 Å) values if required. Click the *Show box* button to update the view.
6. Type in rec.prq into the *Receptor File* box (*see Note 11*) and then click on the *Run cGRILL* button to evaluate, at each grid point, the interaction energy between the whole receptor and five different probes combining van der Waals (Lennard–Jones potential), electrostatic (Coulombic), and hydrogen bonding [19] (geometry-based) terms (*see Note 12*).
7. In the cGRILL plugin window click on the *Load Grids* tab and type in the name of the directory containing the results from either the current session or a previously saved analysis.
8. Display the calculated affinity maps (CH<sub>3</sub>, NH<sub>4</sub>, O, OH, and Hydrophobic) and fine-tune the contours by wisely changing the default cutoff energy values (-1.0 kcal mol<sup>-1</sup>). Display waters in object 2UW3 and note the very precise identification of many hydration sites by the OH, O, and NH<sub>4</sub> probes (Fig. 1).
9. Open the hotspots.pdb file in the same PyMOL session and display the hotspots object as dots or spheres (*see Note 13*) to identify the most favorable regions for interaction with hydrophobic groups (white), hydrogen-bond acceptors (red), hydrogen-bond donors (blue), and mixed hydrogen bond accepting/donating hydroxyl groups (yellow).
10. Note the three close hydrophobic hotspots laid over the phenyl ring of the base fragment, as well as the good superposition



**Fig. 1** PyMOL Viewer window displaying Protein Kinase B as a pink cartoon with a bound inhibitor, as found in PDB entry 2uw7 [12], in stick representation (*left*) and main menu of the cGRIFFL GUI (*right*). The calculated energy contours for hydrophobic, O, NH<sub>4</sub>, and OH probes are shown as wireframe mesh in magenta, red, blue, and cyan colors, respectively. Coalescence of these maps into hotspots (*dotted spheres*) provides precise locations for the placement of fragments possessing atoms with specific hydrogen bonding or hydrophobic properties. Note the good agreement found between hydrophilic hotspots and crystallographic water molecules (*cyan cross marks*), as well as the good overlay of NH<sub>4</sub> and hydrophobic contours onto the piperidine and 4-chlorophenyl substituents, respectively, attached to the 4-phenyl-1*H*-pyrazole base fragment

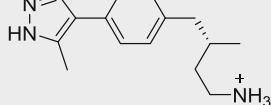
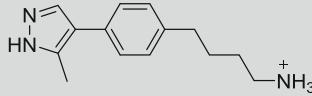
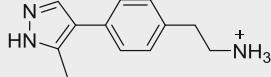
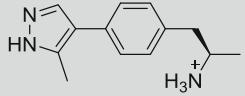
of two mixed hydrogen bond donor/acceptor hotspots onto the two pyrazole ring nitrogens (Fig. 1).

11. Also note (a) the mixed hydrogen bond donor/acceptor hotspots that spread along the bottom of the pocket pointing towards an adjacent cavity and (b) the positively charged cluster with a mixed hydrogen bond donor/acceptor character located near the DFG motif, a key element in the process of activation/inactivation of many kinases [20].

### 3.4 Placement and Linking of New Fragments

1. Use *linker.py* to explore every possible combination resulting from attaching each fragment contained in the different libraries to the base fragment at the defined position (e.g., the carbon at the *para* position of the phenyl ring labeled as C11 in object *ligH*) (*see Note 14*):  
`linker.py ligH.pdb C11 <path_to/fragment_library>`
2. Use *scorer.py* to compute the MW and the PSA of the candidates (*see Note 15*) by means of the OpenBabel Python wrapper *pybel* [21] and to evaluate SEI and BEI for all those compounds with scores that are compatible with favorable binding free energies:

**Table 1**  
**Chemical structures and BEI and SEI values of the top scoring molecules resulting from the first optimization round in the search for PKB inhibitors**

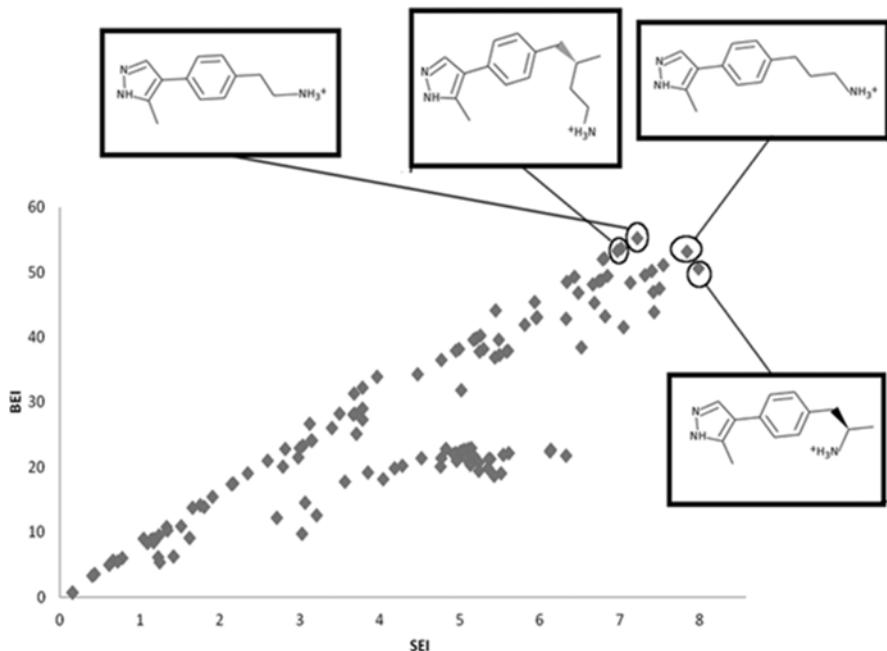
Compound	BEI	SEI
	55.2	7.2
	53.1	7.8
	55.1	7.5
	50.5	8.0

```
scorer.py rec.prq <path_to_dir/where/candidates_
will_be_stored>
```

- Import the CSV file that is generated by the program into a standard worksheet to produce a table containing the scores, LEIs, and compound names (Table 1) and plot the results for visual inspection (Fig. 2).
- Visualize the docking poses of the candidates (pdb.sol files) in PyMOL.

#### 4 Notes

- The software described in this protocol is available for downloading from <http://farmamol.uah.es/soft/fragments/> under the open source license GPL v2.
- Depending on their source, some fragment libraries may require some previous processing. 3D coordinates, if missing because the library consists only of SMILES strings [22] or 2D graphs, can be generated using the OpenBabel *obgen* command or CORINA software [23]. Input files (target, ligand and fragments) must be in Protein Data Bank (PDB) format. In addition, for the target receptor, radii (*R*) and atom point charges (*Q*) are required in the last two columns (PRQ or “inverse” PQR format) in place of the default occupancy and *B*-factor values, respectively, for all atoms. See also Note 8.



**Fig. 2** The BEI vs. SEI efficiency plane for the putative Protein Kinase B inhibitors resulting from attaching different amine-containing fragments to the base fragment. The most promising compounds from the first round which will be used in the next fragment linking step have been *highlighted*

3. cGRILL can be used in command-line mode or within the molecular visualization and editing program PyMOL [11] as a GUI plugin.
4. MM-ISMSA is an ultrafast and accurate force field-based scoring function that comprises (a) a molecular mechanics (MM) part that relies on a 12–6 Lennard–Jones potential; (b) an electrostatic component calculated by means of an implicit solvent model (ISM) [24] that includes individual desolvation penalties for each partner in the protein–ligand/fragment complex; and (c) an SA contribution that accounts for the loss of water contacts upon protein–ligand/fragment complex formation. Since force field-based scoring functions are known to be well suited for pose prediction in docking and to discriminate efficiently amongst native and nonnative candidates [25], MM-ISMSA is the default scoring function for sampling and final evaluation in our protocol.
5. The empirical function ChemScore decomposes the binding energy in terms of (a) a lipophilic contribution (only for non-polar atoms), (b) hydrogen bonding interactions (with a geometry-dependent function), (c) metal interactions (when present and only for hydrogen bond acceptor atoms), and (d) an entropic penalty for the freezing of any rotatable bond

during the binding event (proportional to the number of frozen rotatable bonds). In our implementation, this function lacks any penalty terms for highly strained ligand conformations or for the atomic clashes sometimes observed for very tight binding molecules in X-ray crystal structures. Therefore it will perform optimally in the evaluation of final poses that are force field-compliant in terms of both geometries and energies.

6. The HYDE (HYdration and DEhydration) scoring function assumes that the main contributions to the binding free energy arise from hydrogen bonding interactions between the target and the ligand/fragment and also that the accompanying desolvation event can either favor or penalize binding depending on the nature of the interacting chemical groups. The hydrophobic/hydrophilic nature of the atoms is determined by means of  $\log P$  atomic contributions using empirically deduced coefficients from experimental values. The free energy is estimated as a sum of the goodness of the hydrogen bonds' geometry and the change in solvent-accessible SA of both fragment and target upon complex formation [14, 26, 27].
7. To simultaneously optimize both indices, the sum of their squared values is computed. This information will help the user to decide which of the best possible candidates will be selected as the new scaffold for the next round of growing.
8. If the PDB2PQR tool [28] or the H++ (<http://biophysics.cs.vt.edu/>) or PROPKA (<http://propka.ki.ku.dk/>) web interfaces are used the standard output file in PQR format can be converted to the corresponding PRQ file by using the *autorec.py* script included in the software distribution. There is no need to this, however, if the PyMOL GUI is used because cGRILL.py correctly identifies the columns containing charges and radii irrespective of their relative position.
9. An experimentally confirmed fragment hit at the starting location (as shown in this proof-of-concept example) is not necessary but it increases the odds of a successful final design. Besides, and depending on the structure–activity landscape of the target, it is possible that the resulting optimized compounds will not share the binding mode of the starting fragment [29]. This risk can be minimized by using feature-rich fragments that establish relatively strong and defined interactions with the target. On the other hand, docking-based poses are prone to very well-known errors [30] and do not always ensure that the final molecule will interact with the target as predicted.
10. <http://openbabel.org/> [31].

11. Protein atoms are internally characterized according to their connectivity (bond order), ring state, chemical type, and non-bonded parameters using the generalized AMBER force field (GAFF) [32].
12. These probes are thought to summarize the main stereo-electronic properties of the binding pocket and are defined as follows: lipophilic ( $\text{CH}_3$ ), hydrogen bond donor ( $\text{H}_4\text{N}^+$ ), hydrogen bond acceptor ( $=\text{O}$ ), mixed hydrogen bond donor/acceptor ( $-\text{OH}$ ), and hydrophobic. cGRILL implements the extended atom concept to simplify the probes and to speed up the calculations [13]. The probes are reduced to their central atom with its partial charge increased depending on the atoms attached to it. Accordingly, the hydrogen bond acceptor probe has an assigned charge of  $-0.37e$  to better represent the partial negative potential on the oxygen atom when it is in a carbonyl group. On the contrary, the lipophilic C atom probe is neutral (charge  $0e$ ) and therefore only van der Waals interactions are calculated for it. The functions for hydrogen bond donor, acceptor and mixed donor/acceptor probes include an extra term (besides van der Waals and electrostatic energies) that accounts for the geometry of the hydrogen bond and depends on: (a) the distance between acceptor and hydrogen atoms, (b) the angle between donor, hydrogen and acceptor atoms, and (c) the relative orientation of the planes where the atomic orbitals of the acceptor and the hydrogen atoms are located. The hydrophobic probe is built on the lipophilic one but it adds, as a unique extra feature, the inverse of the default hydrogen bonding term. Thus, this probe will identify those regions where the interaction between the receptor and water molecules is unfavorable and the binding of a small molecule (or fragment) is favored by desolvation due to the hydrophobic effect [33].
13. After the mapping of the binding pocket is complete, the program filters out all those grid points for each probe with scores higher than a user-defined cutoff value (the interaction energy, by definition, is negative), which is set by default to  $-12.0$ ,  $-6.0$ ,  $-7.0$ , and  $-1.7$  kcal/mol for  $\text{H}_4\text{N}^+$ ,  $=\text{O}$ ,  $-\text{OH}$ , and hydrophobic probes, respectively. At each of the surviving points the probes compete according to their interaction energy values, and the best of the set becomes the representative probe at this grid point with its associated energy value. These grid points are then coalesced into local minima that are considered interaction “hotspots” and their coordinates (and energy values) are saved for further use. At each grid point the clustering algorithm checks for the energy values of the nearest surviving points (within  $2.0 \text{ \AA}$  of distance) and, if at least one

of these points has a better value than its own, the current grid point is discarded [34]. In the hotspots.pdb file, atoms and residues are named (N, POS); (O, HBA); (S, HBD); and (H, HPH) for coloring purposes so that positively charged hydrogen-bond donating nitrogens, neutral hydrogen-bond accepting oxygens, mixed hydrogen-bond accepting and donating hydroxyls, and hydrophobic atoms, respectively, can be easily identified. Probe-target interaction energy values, on the other hand, can be displayed in PyMOL by using the hotspots' "B-values" as labels.

14. Script *linker.py* takes three different parameters: (a) the name of the file containing the base fragment properly located in the binding site, (b) the name of the heavy atom in the base fragment to which each fragment in the library will be linked, and (c) the path to the fragment library. This heavy atom must have, at least, one hydrogen atom attached to it since this open valence will be used for bonding purposes. This script requires OpenBabel for internal format conversion and automatic assignment of atomic point charges as these are necessary for the next steps.

Sampling is a completely interactive step that is started by choosing the appropriate fragment database(s) depending on the nature of the hotspot(s) closer to the selected point of attachment on the base fragment. Hydrogen atoms are used for connecting both fragments and the bond length is regularized in accordance with the bond type. The SIMPLEX algorithm [35] and GAFF non-bonded terms are then used to optimize the ligand's rotatable bonds and (optionally) the translational and rotational degrees of freedom to fine-tune the pose within the binding pocket and rank the candidates.

15. The sampling and scoring program can be used as a stand-alone optimization tool for other molecules.

---

## Acknowledgement

This work was supported by grants from CICYT (SAF2009-13914-C02-02 to F.G. and SAF2012-39760-C02-02 to F.G. and A.M.) and Comunidad Autónoma de Madrid (S-BIO-0214-2006 [BIPEDD] and S2010-BMD-2457 [BIPEDD-2] to A.M. and F.G.). A.M. acknowledges financial support from Fundación Severo Ochoa through the AMAROUTO program. A.C.C. is the recipient of an FPU grant (ref. AP2009-0203) from the Spanish Ministerio de Educación.

## References

1. Murray CW, Rees DC (2009) The rise of fragment-based drug discovery. *Nat Chem* 1:187–192
2. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent developments in fragment-based drug discovery. *J Med Chem* 51:3661–3680
3. Congreve M, Carr R, Murray C, Jhoti H (2003) A “rule of three” for fragment-based lead discovery? *Drug Discov Today* 8:876–877
4. Jhoti H, Williams G, Rees DC, Murray CW (2013) The ‘rule of three’ for fragment-based drug discovery: where are we now? *Nat Rev Drug Discov* 12:644
5. Reynolds CH, Touinge BA, Bembenek SD (2008) Ligand binding efficiency: trends, physical basis, and implications. *J Med Chem* 51: 2432–2438
6. Abad-Zapatero C (2013) Ligand efficiency indices for drug discovery. Academic Press London
7. Bembenek SD, Touinge BA, Reynolds CH (2009) Ligand efficiency and fragment-based drug discovery. *Drug Discov Today* 14:278–283
8. Abad-Zapatero C, Blasi D (2011) Ligand efficiency indices (LEIs): more than a simple efficiency yardstick. *Mol Informatics* 30:122–132
9. Blasi D, Arsequell G, Valencia G, Nieto J, Planas A, Pinto M et al (2011) Retrospective mapping of SAR data for TTR protein in chemico-biological space using ligand efficiency indices as a guide to drug discovery strategies. *Mol Informatics* 30:161–167
10. Tanaka D, Tsuda Y, Shiyama T, Nishimura T, Chiyo N, Tominaga Y et al (2010) A practical use of ligand efficiency indices out of the fragment-based approach: ligand efficiency-guided lead identification of soluble epoxide hydrolase inhibitors. *J Med Chem* 54:851–857
11. Schrodinger L. PyMOL molecular graphics system, version 1.5. 0.4. See <http://pymol.org>
12. Saxty G, Woodhead SJ, Berdini V, Davies TG, Verdonk ML, Wyatt PG et al (2007) Identification of inhibitors of protein kinase B using fragment-based lead discovery. *J Med Chem* 50:2293–2296
13. Goodford PJ (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 28:849–857
14. Klett J, Nuñez-Salgado A, Dos Santos HG, Cortés-Cabrera Á, Perona A, Gil-Redondo R et al (2012) MM-ISMSA: an ultrafast and accurate scoring function for protein-protein docking. *J Chem Theor Comput* 8:3395–3408
15. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 11:425–445
16. Schneider N, Lange G, Hindle S, Klein R, Rarey M (2013) A consistent description of HYdrogen bond and DEhydration energies in protein-ligand complexes: methods behind the HYDE scoring function. *J Comput Aided Mol Des* 27:15–29
17. Abad-Zapatero C (2007) Ligand efficiency indices for effective drug discovery. *Expert Opin Drug Discov* 2:469–488
18. Gill A, Cleasby A, Jhoti H (2005) The discovery of novel protein kinase inhibitors by using fragment-based high-throughput x-ray crystallography. *ChemBioChem* 6:506–512
19. Boobyer DN, Goodford PJ, McWhinnie PM, Wade RC (1989) New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J Med Chem* 32:1083–1094
20. Treiber DK, Shah NP (2013) Ins and outs of kinase DFG motifs. *Chem Biol* 20:745–746
21. O’Boyle NM, Morley C, Hutchison GR (2008) Pybel: a python wrapper for the OpenBabel cheminformatics toolkit. *Chem Cent J* 2:5
22. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 28:31–36
23. Sadowski J, Gasteiger J, Klebe G (1994) Comparison of automatic three-dimensional model builders using 639 X-ray structures. *J Chem Inf Comput Sci* 34:1000–1008
24. Morreale A, Gil-Redondo R, Ortiz AR (2007) A new implicit solvent model for protein-ligand docking. *Proteins Struct Funct Bioinf* 67:606–616
25. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT et al (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47:1739–1749
26. Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH et al (2006) A critical assessment of docking programs and scoring functions. *J Med Chem* 49:5912–5931
27. Reulecke I, Lange G, Albrecht J, Klein R, Rarey M (2008) Towards an integrated description of hydrogen bonding and dehydration: decreasing false positives in virtual screening with the HYDE scoring function. *ChemMedChem* 3:885–897

28. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G et al (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* 35:W522–W525
29. Brenk R, Naerum L, Gradler U, Gerber HD, Garcia GA, Reuter K et al (2003) Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J Med Chem* 46:1133–1143
30. Cross JB, Thompson DC, Rai BK, Baber JC, Fan KY, Hu Y et al (2009) Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J Chem Inf Model* 49:1455–1474
31. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open babel: an open chemical toolbox. *J Cheminform* 3:33
32. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174
33. Tanford C (1978) The hydrophobic effect and the organization of living matter. *Science* 200:1012–1018
34. Ruppert J, Welch W, Jain AN (1997) Automatic identification and representation of protein binding sites for molecular docking. *Proc Natl Acad Sci U S A* 94:524–533
35. Nelder JA, Mead R (1965) A simplex method for function minimization. *Comput J* 7:308–313

# Chapter 9

## Scoring Functions for Fragment-Based Drug Discovery

Jui-Chih Wang and Jung-Hsin Lin

### Abstract

Fragment-based drug design represents a challenge for computational drug design because almost inevitably fragments will be weak binders to the biomolecular targets of a specific disease, and the performances of the scoring functions for weak binders are usually poorer than those for the stronger binders. This protocol describes how to predict the binding modes and binding affinities of fragments towards their binding partner with our refined AutoDock scoring function incorporating a quantum chemical charge model, namely, the restrained electrostatic potential (RESP) model. This scoring function was calibrated by robust regression analysis and has been demonstrated to perform well for general classes of protein–ligand interactions and for weak binders (with root-mean square of error of about 2.1 kcal/mol).

**Key words** Scoring function, Protein–ligand interactions, Drug design, Fragment-based, Computational, Structure-based, Docking

---

### 1 Introduction

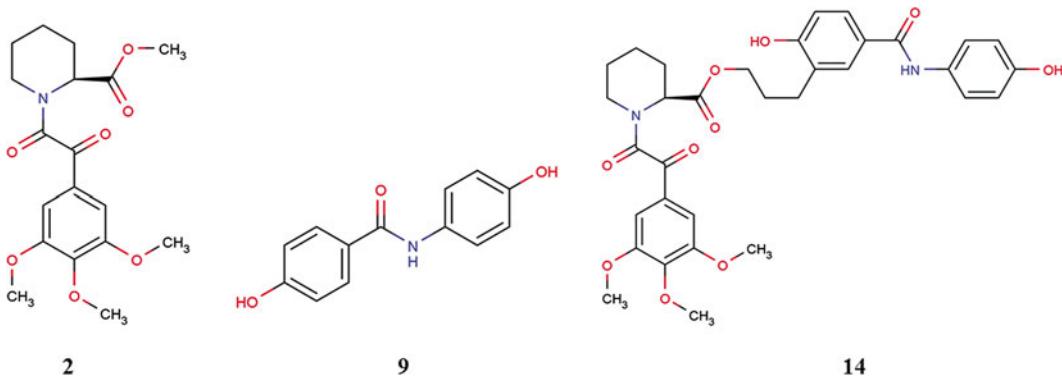
With the approval of vemurafenib by the US Food and Drug Administration on August 17, 2011, fragment-based drug design (FBDD) is no longer just a “promising” approach for drug discovery [1]. However, fragment-based approaches for making leads or drug candidates still represent a challenge for structure-based drug design. Almost inevitably, fragments will be weak binders to the biomolecular targets of a specific disease. It is therefore of crucial importance to be able to discriminate weak binders from non-binders, which usually requires biophysical techniques or computational methods of high accuracy for making the designing process a rational route [2]. Flexibilities of biomolecules have been incorporated in difficult drug design tasks that need to discriminate protein isoforms with high sequence similarity [3]. In the past 2–3 decades, computational approaches for studying the protein–ligand interactions have made a tremendous progress. To reflect the structural variations in the physiological condition, dynamics of biomolecules have been incorporated in computational drug design [4–7].

Effective scoring functions play a central role in the structure-based drug design. We recently refined the AutoDock 4 scoring function [8] with the quantum chemical charge models and robust regression analysis [9]. The refined AutoDock4 scoring functions have been demonstrated to have consistent performance for very large external test sets of protein–ligand complexes [9]. In particular, they also performed very well for weakly interacting protein–ligand complexes [10]. With different binding affinity criteria of weak protein–ligand interactions, our scoring functions showed very consistent performance for three subsets of PDBbind (48 entries for  $K_{i/d} \geq 1$  mM; 158 entries for  $K_{i/d} \geq 100$   $\mu$ M; 335 entries for  $K_{i/d} \geq 10$   $\mu$ M), with the root-mean square of error of about 2.1 kcal/mol [10]. This indicated that these robust scoring functions with the quantum chemical charge models may be suitable for fragment-based drug design, especially when combined with molecular dynamics simulations to accommodate for protein flexibilities.

## 2 Materials

The source of materials and software used in this chapter are mentioned here.

1. The three compounds in the original “SAR by NMR” paper [11] with the optimal binding affinities binding at the first binding pocket, the adjacent second pocket, and both pockets, **2** (2  $\mu$ M), **9** (0.1 mM), and **14** (49 nM), respectively, are used in this chapter. The chemical structures of these three compounds are shown in Fig. 1.
2. The target biomolecule of these three compounds is FK506 binding protein (FKBP). The protein structure is downloaded from the Protein Data Bank [12] and then the protonation is predicted by PDB2PQR [13].



**Fig. 1** The chemical structures of the three compounds in this chapter. The numberings are the same as those in ref. [11]

3. Its partial charges are assigned by the *tleap* program of AmberTools 13 [14].
4. For the ligands, structures and protonations are prepared by MarvinSketch 6.1.0 [15], and their partial charges are calculated by Gaussian 03 [16] and the *antechamber* program of AmberTools 13 [14].
5. The docking jobs are conducted by AutoDock 4.2.5.1 [17], and the input files for docking are prepared by AutoDockTools 4 [17].
6. To use the AutoDock 4 robust scoring functions, one can download the patch files to modify the parameters in the original AutoDock 4 source codes. ([http://jlin.rcas.sinica.edu.tw/~jlin/ScoringScripts/AD4\\_RobustSF.tar](http://jlin.rcas.sinica.edu.tw/~jlin/ScoringScripts/AD4_RobustSF.tar)).

### 3 Methods

#### 3.1 Docking Compound 2 to FKBP

1. Download MarvinSketch (<http://www.chemaxon.com/download/marvin-suite>).
2. Draw the 2D diagram of **2** and predict its protonation state by MarvinSketch, then convert it to the 3D structure with explicit hydrogens added.
3. Optimize the ligand conformation and conduct quantum chemical calculation with Gaussian 03 (<http://www.gaussian.com>) at the HF/6-31G\* level. The Gaussian script for this task, *lig\_2.g03.in*, is as follows (only partial coordinates are shown):

```
--Link1--
%chk=lig_2.g03
%Mem=512MB
# opt HF/3-21g Test
lig_2.pdb.g03
```

0 1			
C1	0.942	3.137	-2.486
C2	-3.229	-2.939	1.280
C3	0.420	4.559	-2.824
...			
O47	-1.345	1.262	-0.900
O48	-2.005	-3.660	1.215
O49	2.160	-1.952	3.231

```
--Link1--
%chk=lig_2.g03
%Mem=512MB
```

```
# sp HF/6-31g* Test SCF=Tight Pop=MK IOp(6/33=2)
iop(6/42=6) geom=AllCheck
```

The command to run Gaussian 03 reads:

```
%g03<lig_2.g03.in>lig_2.g03.out
```

4. Download AmberTools 13 (<http://ambermd.org/#AmberTools>). The RESP partial charges can then be calculated by running the *antechamber* program of AmberTools 13:
 

```
% antechamber -i lig_2.g03.out -fi gout -o lig_2.mol2 -fo mol2 -c resp -at sybyl
```
5. Prepare the ligand pdbqt file with RESP charges by the python script *prepare\_ligand4.py* in AutoDockTools (<http://autodock.scripps.edu/resources/adt>):
 

```
% prepare_ligand4.py -l lig_2.mol2 -C -v -o lig_2.pdbqt
```
6. Download the FKBP structure (1D06) from the Protein Data Bank (<http://dx.doi.org/10.2210/pdb1d06/pdb>). Predict the protonation of the protein by the PDB2PQR server ([http://nbcr-222.ucsd.edu/pdb2pqr\\_1.8/](http://nbcr-222.ucsd.edu/pdb2pqr_1.8/)).
7. Assign Amber Parm99SB charges for the receptor by *tleap* of the AmberTools 13. The commands for *tleap* are shown as follows:
 

```
>1d6o=loadpdb 1d6o-protonated.pdb
> saveamberparm 1d6o 1d6o.top 1d6o.crd
> savepdb 1d6o 1d6o_exleap.pdb
> quit
```
8. There are many possible ways to transfer Amber parm99SB charges into the pdbqt file. Here we present an automatic procedure where the pqr file is used to deposit charges, and then the conversion from the pqr file to the pdbqt file is done by *prepare\_receptor4.py*. The perl scripts *rename\_pqr.pl* and *replace\_charge\_pdbqt.pl* are also needed to map atoms between different file formats, which can be downloaded at [http://jlin.rcas.sinica.edu.tw/~jlin/ScoringScripts/rename\\_pqr.pl](http://jlin.rcas.sinica.edu.tw/~jlin/ScoringScripts/rename_pqr.pl) and [http://jlin.rcas.sinica.edu.tw/~jlin/ScoringScripts/replace\\_charge\\_pdbqt.pl](http://jlin.rcas.sinica.edu.tw/~jlin/ScoringScripts/replace_charge_pdbqt.pl), respectively.
 

```
% ambpdb -p 1d6o.top -pqr<1d6o.crd>rec.pqr
% perl rename_pqr.pl 1d6o_exleap.pdb rec.pqr>rec_rename.pqr
% prepare_receptor4.py -r 1d6o_exleap.pdb -A 'None' -U 'None' -v -o rec_leap.pdbqt
% perl replace_charge_pdbqt.pl rec_rename.pqr rec_leap.pdbqt>receptor.pdbqt
```
9. Prepare the gpf file (the grid maps parameter file) for the input of AutoGrid and the dpf file (the docking parameter file) for the input of AutoDock by AutoDockTools with the default settings.
 

```
% prepare_gpf4.py -l ligand.pdbqt -r receptor.pdbqt
% prepare_dpf4.py -l ligand.pdbqt -r receptor.pdbqt
```

10. Modify the AutoDock parameter file (for using our refined scoring function), the number of grid points, and the grid center in the gpf file and the dpf file:

*receptor.newgpf*:

```
parameter_file AD4.1_RP_wN_110511.dat
npts 26 22 24
gridfld receptor.maps.fld      # grid_data_file
spacing 0.375                  # spacing(A)
receptor_types A C HD N NA OA SA  # receptor atom types
ligand_types A C N OA           # ligand atom types
receptor receptor.pdbqt        # macromolecule
gridcenter 57.311 -4.091 -0.814
smooth 0.5                      # store minimum energy w/in
rad(A)
map receptor.A.map              # atom-specific affinity map
map receptor.C.map              # atom-specific affinity map
map receptor.N.map              # atom-specific affinity map
map receptor.OA.map             # atom-specific
affinity map
elecmap receptor.e.map          # electrostatic
potential map
dsolvmap receptor.d.map         # desolvation
potential map
dielectric -0.1465              # <0, AD4 distance-dep.diel;>0,
constant
ligand_receptor.newdpf:
parameter_file AD4.1_RP_wN_110511.dat
autodock_parameter_version 4.2    # used by autodock to
validate parameter set
outlev 1                         # diagnostic output level
intelec                          # calculate internal electrostatics
seed pid time                     # seeds for random generator
ligand_types A C N OA             # atoms types in ligand
fld receptor.maps.fld            # grid_data_file
map receptor.A.map                # atom-specific affinity map
map receptor.C.map                # atom-specific affinity map
map receptor.N.map                # atom-specific affinity map
map receptor.OA.map               # atom-specific
affinity map
elecmap receptor.e.map            # electrostatics map
desolvmap receptor.d.map          # desolvation map
move ligand.pdbqt                 # small molecule
about -0.0157 -0.027 -0.0478    # small molecule center
```

```

tran0 random # initial coordinates/A or
random
axisangle0 random # initial orientation
dihe0 random # initial dihedrals (relative) or
random
tstep 2.0 # translation step/A
qstep 50.0 # quaternion step/deg
dstep 50.0 # torsion step/deg
torsdof 7 # torsional degrees of freedom
rmstol 1.0 # cluster_tolerance/A
extnrg 1000.0 # external grid energy
e0max 0.0 10000 # max initial energy; max num-
ber of retries
ga_pop_size 150
ga_num_evals 25000000
ga_num_generations 27000
ga_elitism 1 # number of top individuals to
survive to next generation
ga_mutation_rate 0.02 # rate of gene mutation
ga_crossover_rate 0.8 # rate of crossover
ga_window_size 10 #
ga_cauchy_alpha 0.0 # Alpha parameter of Cauchy
distribution
ga_cauchy_beta 1.0 # Beta parameter Cauchy
distribution
set_ga # set the above parameters for GA
or LGA
sw_max_its 600
sw_max_succ 4 # consecutive successes before
changing rho
sw_max_fail 4 # consecutive failures before
changing rho
sw_rho 1.0 # size of local search space to
sample
sw_lb_rho 0.01 # lower bound on rho
ls_search_freq 0.06 # probability of performing local
search on individual
set_pswl # set the above pseudo-Solis &
Wets parameters
unbound_model bound
ga_run 10 # do this many hybrid GA-LS runs
analysis # perform a ranked cluster analysis
11. Perform the grid maps calculations and the docking calcula-
tions by AutoDock 4.2.5.1 (see Note 1)
% autogrid4 -p receptor.new.gpf -l grid.glg
% autodock4 -p ligand_receptor.new.dpf -l result.dlg

```

### 3.2 Docking Compound 9 to FKBP

1. Draw the 2D diagram of **9** and predict its protonation state by MarvinSketch, then convert it to the 3D structure with explicit hydrogens added.
2. Optimize the ligand conformation and conduct the quantum chemical calculation with Gaussian 03 at the HF/6-31G\* level. The Gaussian script for this task, `lig_g03.in`, is as follows (only partial coordinates are shown):

```
--Link1--
%chk=lig_9.g03
%Mem=1GB
# opt HF/3-21g Test
lig_9.pdb.g03
```

0 1			
C1	1.923	2.093	0.673
C2	1.705	0.708	0.463
C3	2.796	-0.144	0.155
...			
O26	5.632	2.295	0.149
O27	-5.705	-2.671	0.264
O28	-0.478	1.091	-1.341

```
--Link1--
%chk=lig_9.g03
%Mem=1GB
# sp HF/6-31 g* Test SCF=Tight Pop=MK IOp(6/33=2)
iop(6/42=6) geom=AllCheck
```

3. RESP partial charges can be obtained by the *antechamber* program.  
`% antechamber -i lig_9.g03.out -fi gout -o lig_9.mol2 -fo mol2 -c resp -at sybyl`
4. Prepare the ligand pdbqt file with RESP charges by the python script `prepare_ligand4.py` in AutoDockTools:  
`% prepare_ligand4.py -l lig_9.mol2 -C -v -o lig_9.pdbqt`
- 5-7. Use the same receptor.pdbqt file as the one in “Docking compound 2 to FKBP”.
8. Prepare the gpf file (the grid maps parameter file) for the input of AutoGrid and the dpf file (the docking parameter file) for the input of AutoDock by AutoDockTools with the default settings:  
`% prepare_gpf4.py -l ligand.pdbqt -r receptor.pdbqt`  
`% prepare_dpf4.py -l ligand.pdbqt -r receptor.pdbqt`
9. Modify the number of grid points and grid center in the gpf file and the dpf file.

```

receptor.new.gpf:
parameter_file AD4.1_RP_wN_110511.dat
npts 24 24 24
gridfld receptor.maps.fld      # grid_data_file
spacing 0.375          # spacing(A)
receptor_types A C HD N NA OA SA    # receptor atom
types
ligand_types A C HD N OA          # ligand atom types
receptor receptor.pdbqt        # macromolecule
gridcenter 58.118 -11.946 4.259
smooth 0.5           # store minimum energy w/in
rad(A)
map receptor.A.map      # atom-specific affinity map
map receptor.C.map      # atom-specific
affinity map
map receptor.HD.map     # atom-specific
affinity map
map receptor.N.map     # atom-specific
affinity map
map receptor.OA.map     # atom-specific
affinity map
elecmap receptor.e.map   # electrostatic
potential map
dsolvemap receptor.d.map # desolvation
potential map
dielectric -0.1465       # <0, AD4 distance-dep.diel;>0,
constant
ligand_receptor.new.dpf:
parameter_file AD4.1_RP_wN_110511.dat
autodock_parameter_version 4.2      # used by autodock to
validate parameter set
outlev 1           # diagnostic output level
intelec          # calculate internal electrostatics
seed pid time      # seeds for random generator
ligand_types A C HD N OA      # atoms types in ligand
fld receptor.maps.fld      # grid_data_file
map receptor.A.map      # atom-specific affinity map
map receptor.C.map      # atom-specific
affinity map
map receptor.HD.map     # atom-specific
affinity map
map receptor.N.map     # atom-specific
affinity map
map receptor.OA.map     # atom-specific
affinity map
elecmap receptor.e.map   # electrostatics map
dsolvemap receptor.d.map # desolvation map

```

```

move ligand.pdbqt          # small molecule
about -0.0135 -0.1605 -0.0127      # small molecule
center
tran0 random                # initial coordinates/A or
random
axisangle0 random           # initial orientation
dihel0 random               # initial dihedrals (relative) or
random
tstep 2.0                   # translation step/A
qstep 50.0                  # quaternion step/deg
dstep 50.0                  # torsion step/deg
torsdof 4                   # torsional degrees of freedom
rmstol 1.0                  # cluster_tolerance/A
extnrg 1000.0               # external grid energy
e0max 0.0 10000             # max initial energy; max
number of retries
ga_pop_size 150
ga_num_evals 25000000
ga_num_generations 27000
ga_elitism 1                # number of top individuals to
survive to next generation
ga_mutation_rate 0.02        # rate of gene mutation
ga_crossover_rate 0.8        # rate of crossover
ga_window_size 10            #
ga_cauchy_alpha 0.0          # Alpha parameter of Cauchy
distribution
ga_cauchy_beta 1.0          # Beta parameter Cauchy
distribution
set_ga                      # set the above parameters for GA
or LGA
sw_max_its 600
sw_max_succ 4                # consecutive successes before
changing rho
sw_max_fail 4                # consecutive failures before
changing rho
sw_rho 1.0                   # size of local search space to
sample
sw_lb_rho 0.01
ls_search_freq 0.06
local search on individual
set_pswl
Wets parameters
unbound_model bound
ga_run 10                    # do this many hybrid
GA-LS runs
analysis
analysis
# perform a ranked cluster

```

10. Perform the grid maps calculations and the docking calculations by AutoDock 4.2.5.1 (*see Note 1*)

```
% autogrid4 -p receptor.new.gpf -l grid.glg
% autodock4 -p ligand_receptor.new.dpf -l result.dlg
```

### **3.3 Docking Compound 14 to FKBP**

1. Draw the 2D diagram of **14** and predict its protonation state by MarvinSketch, then convert it to the 3D structure with explicit hydrogens added.

2. Optimize the ligand conformation and conduct the quantum chemical calculation with Gaussian 03 at the HF/6-31G\* level. The Gaussian script for this task, lig\_14.g03.in, is as follows (only partial coordinates are shown):

```
--Link1--
%chk=lig_14.g03
%Mem=1GB
# opt HF/3-21 g Test
lig_14.pdb.g03
```

0 1			
C1	-4.795	1.411	-0.116
C2	-3.261	1.360	0.119
C3	6.953	4.245	4.746
...			
O79	-4.746	-2.020	0.318
O80	-3.940	-0.559	-2.833
O81	5.053	5.198	0.197

```
--Link1--
%chk=lig_14.g03
%Mem=1GB
# sp HF/6-31g* Test SCF=Tight Pop=MK IOp(6/33=2)
iop(6/42=6) geom=AllCheck
```

3. RESP partial charges can be calculated by the *antechamber* program.  
`% antechamber -i lig_14.g03.out -fi gout -o lig_14.mol2 -fo mol2 -c resp -at sybyl`
4. Prepare the ligand pdbqt file with RESP charges by the python script *prepare\_ligand4.py* in AutoDockTools.  
`% prepare_ligand4.py -l lig_14.mol2 -C -v -o lig_14.pdbqt`
- 5–7. Use the same receptor.pdbqt file as “Docking compound 2 to FKBP”.
8. Because the number of rotatable bonds of *lig\_14.pdbqt* is 15, here we use AutoDock Vina 1.1.2 for the subsequent docking,

which is efficient in finding the optimal pose and is able to do multithreading. Next, each docking pose will be rescored by the AutoDock4<sup>RRP</sup> scoring function.

```
% autodock_vina_1_1_2_linux_x86/bin/vina--onfig config.
txt--og vina.out
config.txt:
receptor=receptor_Vina.pdbqt
ligand=ligand.pdbqt
center_x=56.924
center_y=-9.062
center_z=2.224
size_x=15
size_y=15
size_z=11.25
cpu=16
num_modes=10
energy_range=4
exhaustiveness=100
```

9. For rescoring each docking pose from Vina calculation, preparation of dpf is needed. Here, only local minimization is conducted with AutoDock robust scoring function.

```
receptor.newgpf:
parameter_file AD4.1_RP_wN_110511.dat
npts 40 40 30
gridfld receptor.maps.fld      # grid_data_file
spacing 0.375          # spacing(A)
receptor_types A C HD N NA OA SA    # receptor atom
types
ligand_types A C HD N OA          # ligand atom types
receptor receptor.pdbqt        # macromolecule
gridcenter 56.924 -9.062 2.224
smooth 0.5          # store minimum energy w/in
rad(A)
map receptor.A.map      # atom-specific affinity map
map receptor.C.map      # atom-specific
affinity map
map receptor.HD.map      # atom-specific
affinity map
map receptor.N.map      # atom-specific
affinity map
map receptor.OA.map      # atom-specific
affinity map
elecmap receptor.e.map      # electrostatic
potential map
dsolvemap receptor.d.map      # desolvation
potential map
```

```

dielectric -0.1465          # <0, AD4 distance-dep.diel;>0,
constant
pose_I_receptor.new.dpf:
parameter_file AD4.1_RP_wN_110511.dat
autodock_parameter_version 4.2      # used by autodock to
validate parameter set
outlev 1                      # diagnostic output level
intelec                         # calculate internal electrostatics
seed pid time                   # seeds for random generator
ligand_types A C HD N OA        # atoms types in ligand
fld receptor.maps.fld          # grid_data_file
map receptor.A.map              # atom-specific affinity map
map receptor.C.map              # atom-specific
affinity map
map receptor.HD.map             # atom-specific
affinity map
map receptor.N.map              # atom-specific
affinity map
map receptor.OA.map             # atom-specific
affinity map
elecmap receptor.e.map          # electrostatics map
desolvmap receptor.d.map        # desolvation map
move pose_1.pdbqt               # small molecule
about 57.5219 -8.2867 2.3028    # small molecule
center
tran0 57.5219 -8.2867 2.3028    # small molecule
center
tstep 2.0                      # translation step/A
qstep 50.0                      # quaternion step/deg
dstep 50.0                      # torsion step/deg
torsdof 15                      # torsional degrees of freedom
rmstol 2.0                       # cluster_tolerance/A
extnrg 1000.0                    # external grid energy
e0max 0.0 10000                 # max initial energy; max
number of retries
sw_max_its 600
sw_max_succ 4                    # consecutive successes before
changing rho
sw_max_fail 4                    # consecutive failures before
changing rho
sw_rho 1.0                       # size of local search space to
sample
sw_lb_rho 0.01
ls_search_freq 0.06
local search on individual
set_psw1
Wets parameters
unbound_model bound
# set the above pseudo-Solis &

```

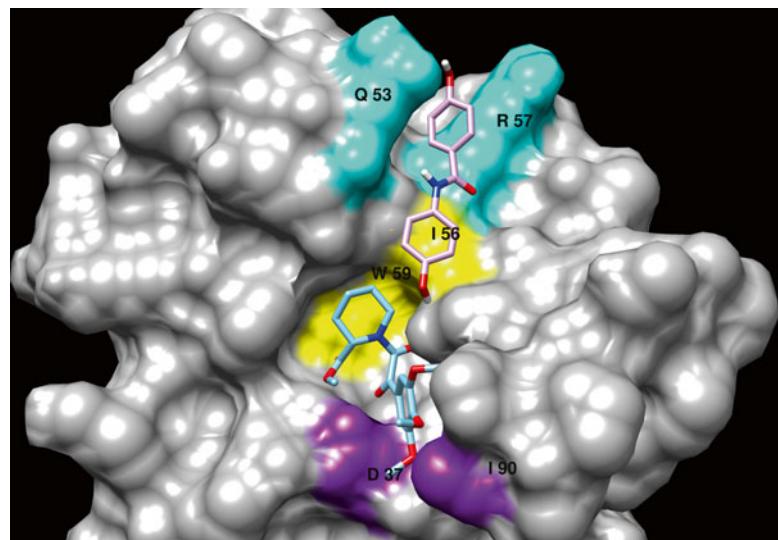
```

do_local_only 10
analysis          # perform a ranked cluster analysis
10. Perform the grid maps calculations and the docking calculations by AutoDock 4.2.5.1 (see Note 2)
    % autogrid4 -p receptor.new.gpf -l grid.glg
    % autodock4 -p pose_1_receptor.new.dpf -l result.1.dlg

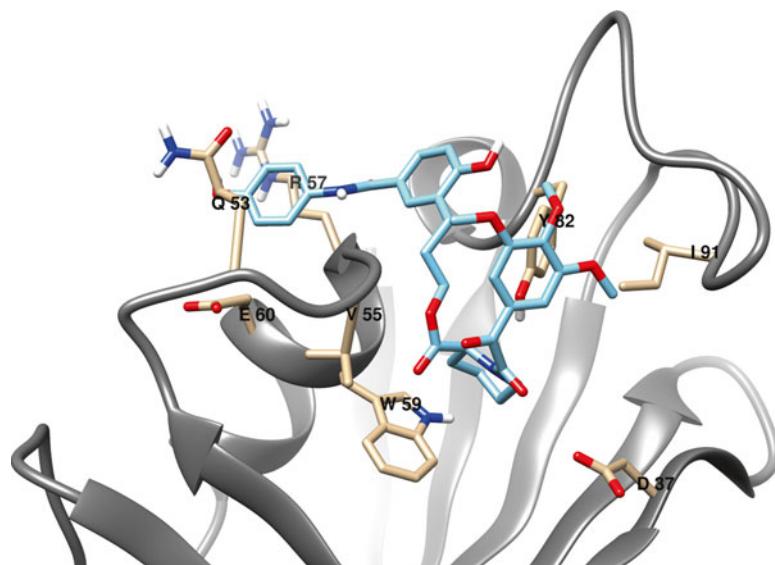
```

#### 4 Notes

- Figure 2 shows the docking poses of **2** and **9** to their target protein, FKBP, which is shown with the molecular surface representation. The binding pose of **2** with third lowest predicted energy (-6.04 kcal/mol) and the binding pose of **9** with lowest predicted energy (-4.22 kcal/mol) are illustrated. It should be noted that the protein flexibility is not yet taken into account in this chapter.
- Figure 3 depicts the binding pose of **14**, the composite compound of **2** and **9**, with lowest predicted energy (-8.12 kcal/mol). The binding pose and contacted residues are consistent with the determination in [11]. From Table 1, it can be seen that the predicted binding affinity is lower than that of



**Fig. 2** FKBP is shown with the molecular surface representation. **2** and **9** are shown with the *stick style*, and with the *cyan* and the *pink color*, respectively. The binding pose of **2** with third lowest predicted energy (-6.04 kcal/mol) and the binding pose of **9** with lowest predicted energy (-4.22 kcal/mol) are illustrated



**Fig. 3** FKBP is shown as *ribbons*. The binding pose of **14** with lowest calculated energy ( $-8.12 \text{ kcal/mol}$ ) is shown with the *stick style* and the *cyan color*. The binding pose and contacted residues are consistent with the determination in [11]

**Table 1**  
**Comparison of the experimental binding affinities and calculated binding affinities of the three compounds**

	<b>2</b>	<b>9</b>	<b>14</b>
Shuker et al.	2 $\mu\text{M}$	0.1 mM	0.049 $\mu\text{M}$
AutoDock4 <sup>RRP</sup>	-6.04 kcal/mol 37.3 $\mu\text{M}$	-4.22 kcal/mol 0.8 mM	-8.12 kcal/mol 1.11 $\mu\text{M}$

the experimental binding affinity, which can likely be attributed to the missing incorporation of protein flexibility in this tutorial.

## Acknowledgements

J.C.W. is supported by the postdoctoral research program of Academia Sinica. Funding from the National Science Council of Taiwan and Research Center for Applied Sciences is greatly acknowledged. We also would like to thank the support from the National Center for High Performance Computing of Taiwan.

## References

1. Bollag G, Tsai J, Zhang J, Zhang C, Ibrahim P, Nolop K, Hirth P (2012) Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat Rev Drug Discov* 11:873–886
2. Mashalidis EH, Sledz P, Lang S, Abell C (2013) A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat Protoc* 8:2309–2324
3. Garcin ED, Arvai AS, Rosenfeld RJ, Kroeger MD, Crane BR, Andersson G, Andrews G, Hamley PJ, Mallinder PR, Nicholls DJ, St-Gallay SA, Tinker AC, Gensmantel NP, Metz A, Cheshire DR, Connolly S, Stuehr DJ, Aberg A, Wallace AV, Tainer JA, Getzoff ED (2008) Anchored plasticity opens doors for selective inhibitor design in nitric oxide synthase. *Nat Chem Biol* 4:700–707
4. Lin JH, Perryman AL, Schames JR, McCammon JA (2002) Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J Am Chem Soc* 124: 5632–5633
5. Lin JH, Perryman AL, Schames JR, McCammon JA (2003) The relaxed complex method: accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers* 68:47–62
6. Amaro RE, Baron R, McCammon JA (2008) An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J Comput Aided Mol Des* 22: 693–705
7. Lin JH (2011) Accommodating protein flexibility for structure-based drug design. *Curr Top Med Chem* 11:171–178
8. Huey R, Morris GM, Olson AJ, Goodsell DS (2007) A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* 28:1145–1152
9. Wang JC, Lin JH, Chen CM, Perryman AL, Olson AJ (2011) Robust scoring functions for protein-ligand interactions with quantum chemical charge models. *J Chem Inf Model* 51:2528–2537
10. Wang JC, Lin JH (2013) Scoring functions for prediction of protein-ligand interactions. *Curr Pharm Des* 19:2174–2182
11. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274: 1531–1534
12. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
13. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* 35(Web Server issue):W522–W525
14. Case TAD DA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Götz AW, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wolf RM, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA (2012) AMBER 13. University of California, San Francisco
15. Marvin 6.1.0 (2013). ChemAxon (<http://www.chemaxon.com>)
16. Frischmijt GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, Revision C.02. Gaussian, Inc, Wallingford, CT
17. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791

# **Part III**

## **Design**

# Chapter 10

## Computational Methods for Fragment-Based Ligand Design: Growing and Linking

Rachelle J. Bienstock

### Abstract

Fragment-based drug design has proved itself as a powerful technique for increasing the sampling and diversity of chemical space and enabling the design of novel leads and compounds. Computational techniques for identifying fragments, binding sites and particularly for linking, growing, and evolving fragments play a significant role in the process. Information from ADME studies and clustering property information in the form of toxicophores and chemotypes can play a significant role in aiding the design of novel, selective fragments with good activity profiles.

**Key words** Fragment-based ligand design, Fragment linking, Fragment growing, Evolving fragments, Fragment libraries, Structure-based drug design, ADME properties and fragment-based design, Chemotypes, Toxicophores

---

### 1 Introduction

Fragment-based ligand design is now well established as an efficient starting point and optimization method in structure-based drug discovery. Pharmaceutical industry-based application of fragment-based screening was first described by Dr. Stephen Fesik and colleagues at Abbott using SAR by NMR [1] with the detection of millimolar binding ligands to  $N^{15}$  labeled protein. Since their initial application, fragment-based methods have shown promise for the identification and development of novel hits and lead compounds. The essential concept behind fragment-based design is the use of a fragment library consisting of weak binders to the identified target. These weak binding fragments (fragment affinities usually in the millimolar range) are then linked and optimized to create appropriate tight-binding drugs. In many cases, experimental methods such as NMR or X-ray crystallography are used to detect fragment binding and initially identify fragments. However, the identification of fragments is often accomplished by computational methods as well.

What are the advantages and why would someone pursue fragment-based design as opposed to traditional virtual screening of a regular ligand library? The fragment-based approached has been pursued because of the advantage for vastly increasing compound chemical diversity. Also screening large diverse libraries, containing complex compounds, incurs the difficulty of problematic functional groups, which must then be removed or modified to increase binding. Traditional high-throughput screening (HTS) library successful hit rates are low (<0.01 %). Well-designed fragment libraries include more diversity than typical large molecule HTS databases; fragment libraries can be designed with lead like compound fragments only and therefore exhibit increased probability of binding to the target of interest. Eighty percent of atoms in fragment hits are retained in the derived lead compound and exhibit less than 1 Å displacement shifts [2]. There has been great success using fragment design with traditional targets, as well as difficult ones like BACE and protein–protein interaction targets.

One advantage of fragment-based drug design is that chemical space can be explored more efficiently by screening collections of small fragments rather than high throughput screening of virtual libraries of larger drug-like molecules. Hann's model for receptor-ligand interactions states that the probability of detecting a hit binding to a target will increase with a smaller less complex fragment. With a library comprising  $10^7$  fragments with 12 heavy atoms rather than  $10^{60}$  drugs with 30 heavy atoms a much greater percentage of fragment space can be screened. It has been estimated that there are  $10^{20}$ – $10^{200}$  molecules with molecular weight (MW) 300–500 Da (i.e., drug-like compounds) more than all the atoms in the universe [3]. It is felt that even a small fragment library will do a better job sampling fragment chemical space than a larger traditional compound library. Fragment-based drug design has a high impact-high hit rate, a high binding efficiency for fragments, and a high proportion of fragment hit atoms directly involved in interacting with the target-binding site. Fragments can easily be optimized. One of the major difficulties encountered with fragment-based drug design is initially identifying fragments, which bind the target since fragments have low affinities for targets (usually on the order of 1 mM or less). Fragments, which bind targets, are identified computationally, or by experimental methods, such as NMR, surface plasmon resonance (SPR), thermal shift, functional screen, X-ray crystallography, isothermal titration calorimetry, or a biophysical assay.

Also some fragments, which bind and seem to be identified, as good fragments in assays are not. They are referred to as “PAINS” (pan-assay interference compounds), have common features (Michael acceptors), and are soft electrophiles that readily react with nucleophilic residues in proteins and bind covalently. However, because they form covalent bonds it makes them unsuitable to use in fragments. Redox cycling compounds are also part of this group [4].

A fragment library for drug design must be chemically diverse and synthetically expandable. Fragments traditionally have low affinity binding (micromolar to millimolar) and high ligand efficiency  $\geq 0.3$  (ratio of free energy of binding to number of heavy atoms). A good fragment library has a range of physicochemical properties, aqueous solubility, molecular diversity, and drug likeness with medicinal chemistry scaffolds. Astek [2] researchers first proposed the “Rule of 3” (Ro3) for fragment design, based on Chris Lipinski’s popular rule of five for general virtual library design. The Rule of 3 for fragment library design is as follows: MW  $\leq 300$ , CLogP  $\leq 3$ , number of H bond donors  $\leq 3$ , number of H bond acceptors  $\leq 3$ . Also usually the number of rotatable bonds is  $\leq 3$  and the polar surface area is  $\leq 60 \text{ \AA}^2$ . This is in contrast to Lipinski’s rule of five, Ro5, for general virtual screening library design proposes MW  $\leq 500$ ; cLogP  $\leq 5$ ; number of H bond donors  $<+5$ ; number of H bond acceptors  $\leq 10$ .

When designing a fragment library, another measure of a ligand quality is the LE (Ligand efficiency):

$$\text{LE} = -\Delta G / \text{HAC} (\text{heavy atom count}) = -RT \ln(\text{IC50}) / \text{HAC}$$

The group efficiency is an estimation of a group’s contribution toward the overall free energy of binding (the heavy atom count (HAC) number in a particular group). The Ligand-lipophilicity efficiency is often thought to be a better measure of successful fragment design since lipophilic molecules have an increased change on binding to any drug pocket (LLE) LLE = pIC50 or pKi-cLogP (or LogD).

Fragment-based drug design is currently such a rapidly developing field that it is often difficult remaining informed. There are two popular blogs which contain a wealth of information regarding meetings, available fragment libraries, publications, and new developments in the field: Dr. Daniel Erlanson’s (Carmot Therapeutics) and Dr. Edward Zartler’s (Quantum Tessera Consulting) Practical Fragments <http://practicalfragments.blogspot.com/> and Dr. Peter W Kenny’s FBDD and Molecular Design (formerly FBDD Literature blog) <http://fbdd-lit.blogspot.com/>.

The 3D Fragment Consortium (<http://www.3DFrag.org>) has recently been established in the UK for not-for-profit drug discovery groups and currently includes as members—University College London, Structural Genomics Consortium, Institute of Cancer Research, University of Cambridge, Peterson institute for Cancer Research, Cancer Research Technology, Cancer Research UK, MRC, and University of Dundee Drug Discovery Unit. The consortium has recently developed a tool 3DFIT = 3D Fragment Idea Tool to determine shape and predicted physicochemical properties of compounds, which is based on commercial software—Accelrys’ Pipeline Pilot tool and ChemAxon’s 3D conformation generator.

---

## 2 Identifying the Fragments and Developing Good Fragment Libraries: Computational Methods

Fragment libraries can be much smaller than typical large molecule virtual screen libraries (1,000–20,000 fragments) [5]. Successful drugs have specific properties—low MW and low lipophilicity. Fragments usually have low ClogP and low MW, are weak in potency, but still form high-quality interactions with the target protein. Fragments overcome entropic barriers to bind to the target and form very high quality interactions. There are some basic requirements for a fragment library—all fragments must have aqueous solubility so that they can be screened at high concentrations; low logP biases a fragment library towards increased solubility.

Computational methods have tried to categorize, classify, and visualize chemical fragment space. Two dimensional (2D) molecular FPS (fingerprints) (MD keys, Daylight FPS, extended connectivity FPs, and Unity 2D FPS), compared using Jaccard, Tanimoto coefficients, principal component analysis, or field similarity methods have been used to compare molecules and are computational methods that can be used to describe diversity in chemical space. Jean-Louis Reymond's group [6] (<http://www.gdb.unibe.ch/>) has designed MQN (multiple quantum numbers) a simple classification system for organic molecules based on only 42 integer value descriptors for molecular structure. Molecules are classified by topological indexes such as atom and ring counts, cyclic and acyclic unsaturations, atoms and bonds in fused rings and electrostatic charges predicted for neutral pH, molecular size and H bond acceptor count, Lipinski's bioavailability rules, Opera lead likeness rules and Congreve fragment likeness rules. Since fragment library composition is critical, computational methods for analysis of the chemical space available for fragments improves and facilitates fragment selection. MQN classifiers can then be used for similarity searches to enrich binding fragments tied to bioactive compounds. This method can also be used to assess diversity of a library.

Jean-Louis Reymond has constructed a database GDB-13 (updated and now called GDB-17) [7, 8] which contains 977 million virtual organic molecules composed of C, N, O, S, Cl for generating a diverse comprehensive fragment library which was then analyzed using his MQN method to subdivide the database into 255 characteristic subsets, and analyzed by principal component analysis using Jsci a publicly available code <http://jsci.sourceforge.net>. A searchable version of the GDB-13 (now GDB-17) database is available at [www.gdb.unibe.ch](http://www.gdb.unibe.ch).

Computational fragment positioning methods such as HSITE, HIPPO, GRID, MCSS, SPROUT, MUSIC, LUDI, and SuperStar have been used for over 20 years as early stage lead optimization techniques. These methods either determine binding site positions

for chemical functional groups based on molecular mechanics potentials, or determine hotspots on the target/receptor. Caveat and HOOK are computational fragment linking first approaches. LUDI was a very early based computational scaffold replacement method [9] for ligand design based on fragments. Newer methods such as Recore, Allegrow, Confirm, Med-Sumo, MOE scaffold replacement, and the CHARLIE/RACHEL Tripos package are used for linking and scaffold replacement.

Some computational fragment design methods have their origins in experimental X-ray techniques, such as the MCSS X-ray methodology pioneered by Drs. Dagmar Ringe, Gregory Petsko [10], and Carla Mattos where crystals were soaked with various organic solutes of small molecular weight to experimentally identify groups, which bound and could be incorporated into ligands, and linked to create tighter binders. SGX's method FAST (Fragments of Active Structures) relies on X-ray crystallography with many library members containing bromine [3] to facilitate detection by crystallographic screening using the X-ray energy tuned to the bromine absorption edge and its chemistry advantage (bromide acts as a leaving group in Suzuki coupling and related reactions).

SILCS (Site Identification by Ligand Competitive Saturation) is a computational *in silico* method [11], based on the experimental MCSS technique to detect multiple binding sites of fragments. Targets are computationally soaked in an aqueous solution of compounds with drug like fragments. Then molecular dynamics (MD) is performed of the target protein in the presence of an aqueous solution of organic solutes. The probability of the fragments binding to different sites on the protein is obtained. The 3D probability distributions of the molecular fragments binding on the protein surface are called “FragMaps” (“hotspots” for binding). The SILCS method quantifies relative ligand affinities by converting the fragmaps into Grid Free Energies. Unrestrained MD simulations can be performed as well to consider target flexibility. With fragments there is significant conformational changes and movement on binding, with movements > 5 Å observed.

GRID interaction, MCSS (multiple copy simultaneous search) and SILCS (site identification by ligand competitive saturation) are computational methods where the interaction between a variety of molecular probes and a receptor are energetically calculated to quantify the low energy-binding pocket. There are other computational fragment mapping programs, such as FTMAP [12] <http://ftmap.bu.edu/login.php>. The FTMAP algorithm searches the protein surface for regions that bind small organic probe molecules using an FFT algorithm, and can be used to identify druggable binding hotspots on targets. These methods are more computationally intensive than other computational methods to geometrically determine the ligand-binding pocket.

Computational design of fragment libraries involves either retrospective analysis (dissection) or forward de Novo design. In retro synthesis, an existing molecule with known activity is fragmented. Lead compounds are cleaved at acyclic bonds. The RECAP method is one of the first computational retro synthetic combinatorial analysis procedures [13] and is implemented in the Open Source publicly available RDKIT <http://www.rdkit.org/docs/api/rdkit.Chem.Recap-module.html>. RECAP performs retro synthesis and can help identify building blocks in active molecules. These building blocks are often called “privileged structures.” RECAP fragments molecules at 11 predefined bond types selected because they are formed by combinatorial synthetic methods. Other widely used fragment methods besides RECAP include Richard Cramer’s Tripos method ChemSpace [14], AllChem Topomer search [15], and Rarey’s FTtrees-FS—feature trees fragment space (available commercially from BioSolveIt) [16]. Eugene Lounkine’s FragFCA [17] method identifies fragments and fragment combinations for compounds having certain activity profiles. Bajorath’s MolBlaster [18] generates fragments based on molecular similarity relationships. There are several other commercial fragmentation methods available—Flux, BioSolveIT’s ReCore, and Chemical Computing Groups MOE fragmentation method. In forward design, fragments are constructed de novo. Vertex’s retro synthetic SHAPES library is an example of a simple organic framework. In silico de novo design and combinatorial libraries are built using prioritizing algorithms for assembly rules, considering molecular attributes and physicochemical properties.

FragVLib [19], (<http://www.unc.edu/~raed/FragVLib.zip>) is a free database mining program for generating a “Fragment-based Virtual Library” using a pocket similarity search of ligand–receptor complexes. Tools that mine databases of ligand–receptor complexes and can generate a library of fragments rely on a graphical representation of interfacial atoms for the ligand–receptor complex. Interfacial atoms are nodes and the distances between them are edges. Pocket similarity matches are performed using a graph type match. The program written in C++ is available as freeware and downloadable. Once fragments are identified another program FragVScreen can be used to search molecules that contain these fragments.

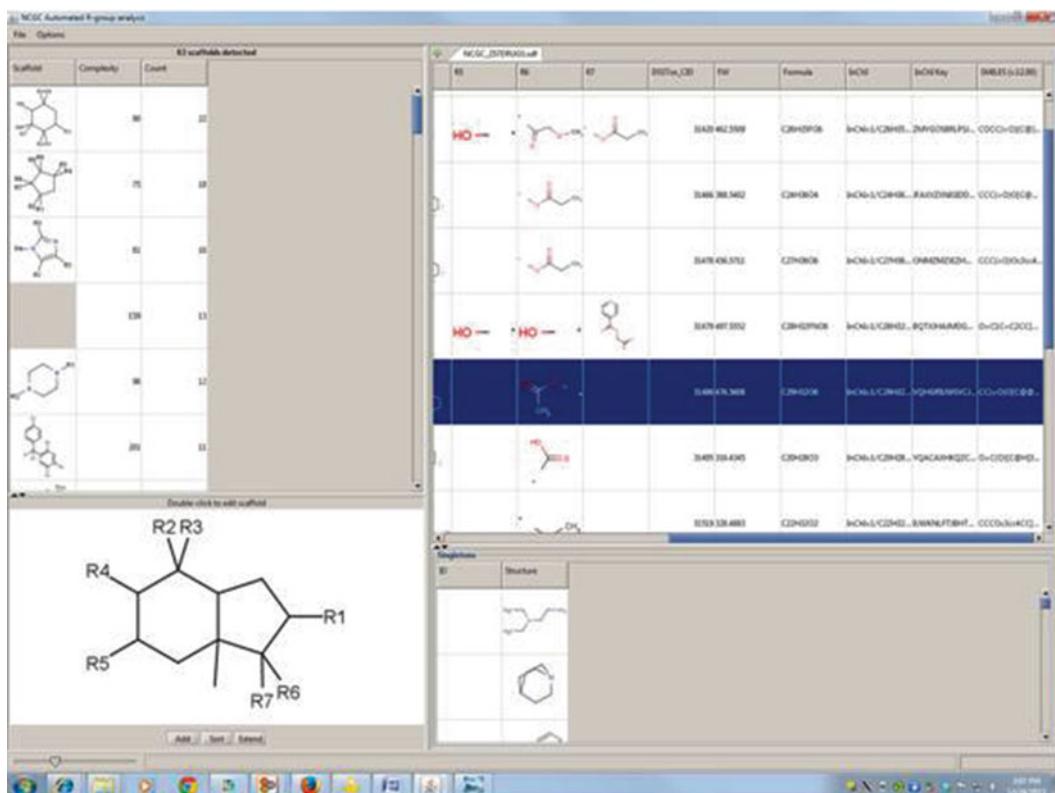
REOS (rapid elimination of swill) (Vertex) is used to filter a fragment database for desirable properties—solubility, MW, lipophilicity, polar surface area, number of rotatable bonds, and H bonding potential. Data mining workflows (Knime, Accelrys Pipeline Pilot, Taverna) are used to screen and filter compounds with substructure filters for toxicophores and other undesirable features and also to screen for halogen enriched fragments to take advantage of halogen bonding (viewed as a plus in ligand design). Workflows can be implemented as well, with diversity selection

algorithms—Similog keys, hole-filling algorithms for diversity selection of fragments. In drug discovery very often a diverse fragment set is paired with a focused fragment set—usually from in silico virtual screening (i.e., computational docking or 3D pharmacophore analysis).

The National Center for Advancing Translational Sciences (NCATS) and National Chemical Genomics Center (NCGC) provides a group of free Java tools, called Tripod (<http://tripod.nih.gov/>) (Therapeutically Relevant Informatics for Prioritization, Optimization, and Development), which perform automated R group analysis and editing of scaffolds, a fragment activity profile, a program, Siponify, for fragment searching, library synthesizer, scaffold activity diagram and scaffold hopper. Figure 1 illustrates the graphical interface for the R group analysis.

## 2.1 Some Newer Methods Applied to Fragment Design

Fragment-based chemogenomics adds genomic information to inform the chemistry. It is a new method, which uses information from protein–ligand binding sites of genetically related protein family members to search all related proteins with the libraries of



**Fig. 1** This image illustrates the graphical interface for the R group analysis performed using the free Java tools, called Tripod (<http://tripod.nih.gov/>) developed by NCATS and NCGC which perform automated R group analysis

small fragment like molecules. One example where this method was successfully applied has been to a family of 4 G-coupled receptor and related adrenaline receptors [20].

CANVAS HF (hole filling) [21] a new commercial method developed by Schrödinger, uses distances computed from 2D fingerprints to fill regions in a chemical library which are not represented, in order to add additional chemical diversity. This software uses a greedy selection algorithm, combined with simulated annealing to minimize nearest neighbor 2D fingerprint similarities among structures selected with respect to an existing compound library.

Virtual fragment screening using molecular fingerprints derived from fragment affinity data is a novel method called FLAP (Fingerprints of Ligands and Proteins). FLAP uses four point pharmacophores derived from molecular interaction fields to align molecules (GRID Molecular Interaction Fields) [http://www.moldiscovery.com/soft\\_flap.php](http://www.moldiscovery.com/soft_flap.php) [22].

SERAPHiC is a fragment–protein dataset available developed from selected fragment–protein complexes. This is a test set which can be used for in silico protocol assessment and software development <http://www.iit.it/en/drug-discovery-and-development/seraphic.html> [23].

Fragment-based shape descriptors can be generated using a Connolly rolling probe sphere. An updated triangular form is used (<https://www.artemisdiscovery.com/>) for a fragment-based shape signature implementation method. Shape signatures can be used to screen databases and can produce hits active against a target. Since the method does not involve chemical structure, it allows for identification of novel chemical classes and modulators of drug targets and performs novel scaffold hopping. This method has been used to compute shape signatures for the ZINC database [24].

S4MPLE [25] (Sampler For Multiple Protein–Ligand Entities), along with two Java programs, Genlinkers DB and JmolEvolve, (based on the ChemAxon API) is a fragment-based ligand design method that performs conformational sampling using a genetic algorithm and suggests fragment growth using bridging waters. S4MPLE (x86\_64) can be uploaded from <http://infochim.u-strasbg.fr>; however, a license is needed for the ChemAxon (commercial software) growth/linking tools. GenLinkers and JMolEvolve “evolve” compounds by starting with lead like and drug like fragments to create a library. GenLinkers creates linkers using RECAP-like rules and methods and JMolEvolve generates new molecules by combining the fragments in the database with the GenLink generated linkages. Many useful software tools for fragment-based design are available and are compiled on websites, as shown in Table 1.

**Table 1**

**Many useful software tools for fragment-based design are available and are compiled on these websites**

Description	Website
Bruno Villoutreix's group has compiled a list of useful computational fragment-based drug design tools	<a href="http://www.vls3d.cm/links.html">http://www.vls3d.cm/links.html</a>
O'Boyle, Linux4Chemistry	<a href="http://www.redbrick.dcu.ie/~noel/linux4chemistry">http://www.redbrick.dcu.ie/~noel/linux4chemistry</a> ; <a href="http://www.linux4chemistry.info/">http://www.linux4chemistry.info/</a>
Drs Zoete and Grosdidier, Click2Drug	<a href="http://www.click2drug.org/">http://www.click2drug.org/</a>
Dr Sung Kwang Lee	<a href="http://www.qspr.pe.kr/my/index.php">http://www.qspr.pe.kr/my/index.php</a>
Etox Library	<a href="http://cadd.imim.es/etox-library">http://cadd.imim.es/etox-library</a>

### 3 Growing, Linking and Evolving Fragments

Once candidate fragments are identified from a library they are merged or linked together, adding functionalities to increase binding and selectivity. When linking fragments the LE (ligand efficiency) and affinity should be greater than the sum of binding free energy of the two-parent fragments—referred to as the “super-additivity” of fragment binding energies. Computational fragment linking is very difficult and often what is proposed from computational studies is not synthetically feasible. Privileged-fragment merging involves recognizing fragments inside larger molecules, pharmacophore modeling and overlays of X-ray structures. Merging of fragments identified through screening can be performed with cocrystal structures of multiple fragments overlaid. It is important when merging fragments not to make major changes to the core scaffold.

The binding modes of the parent fragments must be maintained on linking. SPROUT [26] (<http://www.simbiosys.com/sprout/overview.html>), marketed by Symbiosis, uses steric electrostatic H bonds, dispersion or van der Waals, and hydrophobic interactions as constraints when adjoining templates. Other common commercial software for designing linkers between two fragments includes CCG MOE, BioSolveIt Recore, and ChemAxon Reactor. GANDI [27] (Genetic Algorithm-based de Novo Design of Inhibitors) is a fragment-based method that generates leads by joining predocked fragments with linkers using a parallel genetic algorithm. Predocked fragments are encoded by the genetic algorithm, and linker fragments are evaluated using lookup tables (tabu search). GANDI (<http://www.biochem-caflisch.uzh.ch/download/>) is available for download free to nonprofits.

Fragment evolution and optimization can be built around a well-characterized target or existing SAR (Structure–Activity Relationship), or 3D QSAR (CoMFA and CoMSIA). A selected fragment can be grown by doing a QSAR search and picking molecules similar to the identified fragment. The SAR results for two sites are applied for fragment screening. Suitable fragments obtained were added to a scaffold, and then docked and the activity predicted by the 3D CoMSIA model [28]. Docked poses are then compared to crystal structures of the fragment with the target. The fragment-linking step can involve several different types of chemistry.

Autogrow [29] from MacCammon’s group <http://autogrow.ucsd.edu/> is a free open source software, which optimizes candidate ligands using rules of “click chemistry.” The program grows fragments using an evolutionary algorithm and discards non-drug like candidates. Autogrow 3.0 contains autoClick Chem [30] reactions (<http://autoclickchem.ucsd.edu>) programmed into it so that unlike previous versions designed molecules from fragments are synthesizable in a practical way. The program has Lipinski’s rules embedded so it only designs drug-like molecules.

BREED [31] is a computational method for fragment merging which is widely employed. SPLICE and BREED mix and match sets of overlaid 3D X-ray structures using combinatorics. BREED uses “Fragment Shuffling” and aligns 3D coordinates of two ligands and recombines the fragments or substructures at overlapping bonds to generate new hybrid molecules. MED-hybridize and FLUX are other 3D ligand-based methods. MED-hybridize relies on ligand structural information from the PDB (Protein Databank). FLUX uses a stochastic search algorithm for combinatorics. LigMerge [32] is an automated algorithm for swapping chemical substituents in known ligands to generate new ligands. It identifies common substructures in ligands and superimposes two substructures and then scrambles them at each atom to generate multiple compounds related to the known inhibitors. This is a stand-alone ligand-based tool that does not require knowledge of the receptor structure unlike BREED, and is available for download (<http://www.nbcr.net/ligmerge>).

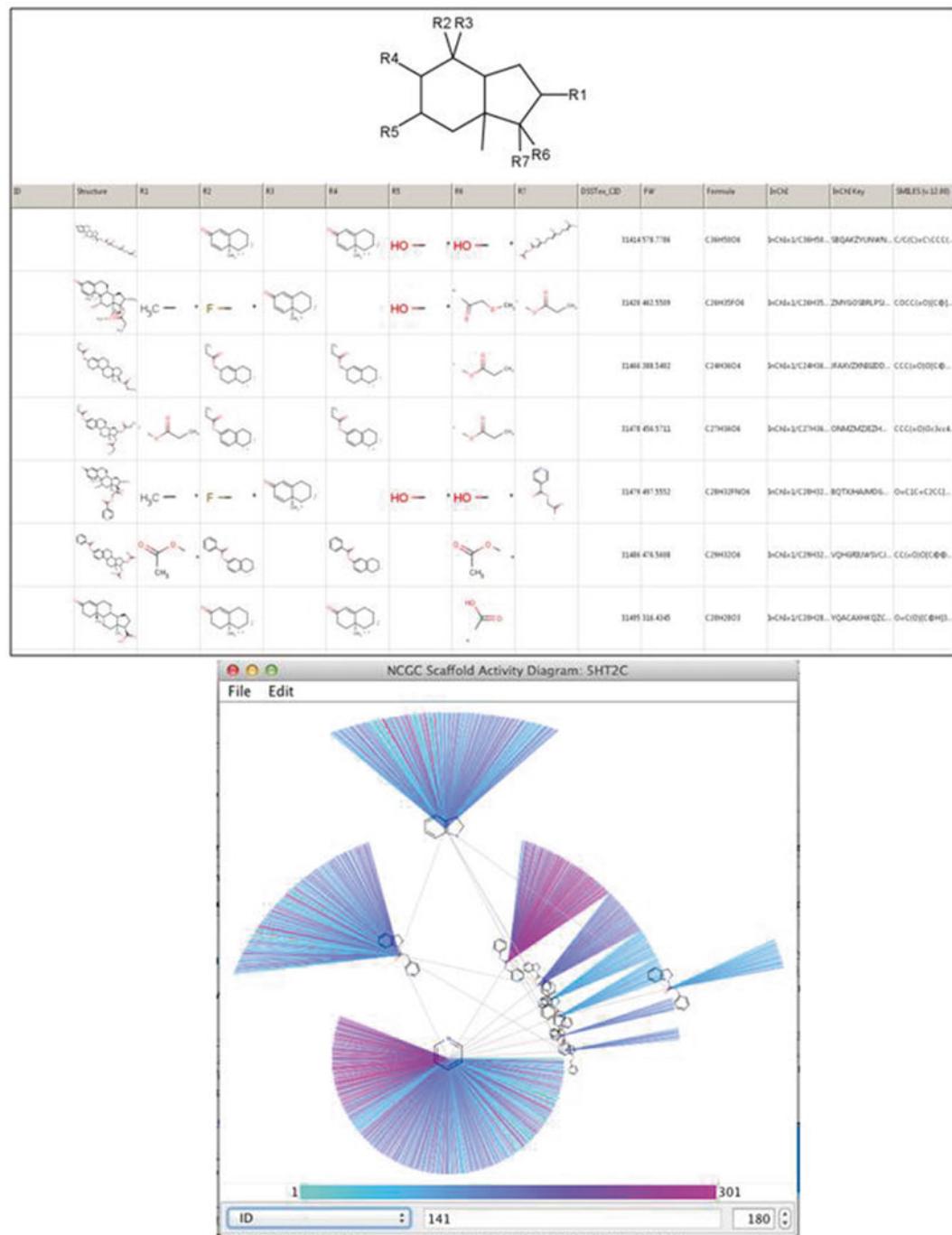
Fragment tethering—the formation of a disulfide bond between the fragment and a cysteine residue in the target protein is an additional novel method. Ideally linkers would not perturb the optimal binding geometry of the fragments and not have conformational flexibility so that they would not increase the entropy of binding. Free energies of binding shows that flexibility and linker strain can have an impact on binding affinity [33]. A new approach uses a piece of the full substrate fragment that is still binding competitively, and the substrate fragment is modified with a chemical “handle” to connect via variable length linker to a random fragment library.

CONFIRM (connecting fragments found in receptor molecules) (Wyeth) [34] is a pre-prepared library of linkers searched to match criterion found from experimental or computational studies about fragments within the target binding site. Other computational tools available for growing fragments include; SPROUT, LEGEND, LUDI, GROWMOL, LigBuilder, SkelGen, and SMoG. Techniques, which favor the linking approach vs. growing approach, are LUDI, HOOK, PRO\_LiGAND, LigBuilder, SPLICE/RACHEL, and CAVEAT. Linking fragments positions them in the target-binding site and connects them to each other by designed linkers to construct a compound that satisfies significant interactions with the target. Many of these methods search databases of bridges using substrate pattern search functions, with a given set of distances between attachment points and atoms used to form the query to search the linker database. The linkers found are than joined using a combinatorial method. Once fragments are linked, computational docking of molecules is performed, computing strain energy of docked poses.

ReCore [35] is a popular commercial software (BioSolveIT), which performs scaffold hopping, and fragment linking (<http://www.biosolveit.de/ReCore/index.html?ct=1>). The Recore approach is to use a database of 3D molecular structures converted into a fragment database. Given the geometric arrangement of two or more exit vectors and a pharmacophore feature, the algorithm finds 3D fragments within the database, generated by combinatorial analysis of 3D compound libraries, fulfilling all constraints. The fragmentation method is similar to RECAP, in terms of the rules considered for generating fragments-no cut points within ring fragments, with rules usually cutting C-C, C-O, C-N, and C-H bonds in specific environments.

Scaffold hopping or lead hopping where the central part of known active molecule is varied instead of an R group is often used to improve ADME properties, and avoids structures with strained conformations. Programs which perform scaffold hopping include CAVEAT, BROOD, sparkV10, Core Hopping, SHOP, Scaffold Replacement, PraFrag, and the OEChem toolkit C++ Web application [36]. CAVEAT was one of the first and premiere programs designed to do scaffold hopping and is licensed and available through Paul Bartlett's group at UC Berkeley (<http://www.cchem.berkeley.edu/pabgrp/Data/caveat.html>). CAVEAT searches a database of geometric relationships using bond pairs.

NCGC developed Scaffold Hopper, a freely available tool for automated R-group analysis (Fig. 2a is an illustration of the graphical interface for the Scaffold hopper as well as graphical visualization tools (Fig. 2b) available for scaffold activity analysis). It is a self-contained Java web start application. <https://tripod.nih.gov/ws/hopper/hopper.jnlp>. The software generates “reasonable” R-group tables for a given dataset and performs Scaffold-based



**Fig. 2 (a)** This is an illustration of the graphical interface for the NCGC Scaffold hopper, a freely available tool for automated R-group analysis (<https://tripod.nih.gov/ws/hopper/hopper.jnlp>) as well as graphical visualization tools **(b)** available for scaffold activity analysis. (sad.jnlp) is downloadable from <http://tripod.nih.gov> and can perform a network oriented view of fragment/scaffold collections

“clustering” of the data. The NCGC scaffold activity diagram (sad.jnlp) is downloadable from <http://tripod.nih.gov> can perform a network oriented view of fragment/scaffold collections.

The SAMPL3 Challenge [37] was a public competitive challenge initiated to evaluate performance of fragment-based design methodology. Statistical Assessment of the modeling challenge at (<http://sampl.eyesopen.com/>) method-input summarizes how incorporation of different modeling techniques impacts the prediction accuracy including protein dielectric models and charged vs neutral ligands. The linker problem is a significant one in fragment-based ligand design. Linking close fragments depends on bond length and angle strain and high-energy conformations can be created and electronic properties change. Simulated annealing of rigid fragments does not answer these problems, and new techniques, such as constrained fragment annealing (CFS) must be performed.

---

## 4 ADME Properties and Fragment-Based Design

Fragments from large collections can also be used to generate models for QSAR studies, and as filters and scaffolds, and to explore chemotypes. Using fragment libraries to predict drug toxicity is a way to minimize expensive drug failures and to identify potential toxicity early in the drug discovery process. An extension of the use of QSAR methods for toxicity prediction is the use of fragments for predicting lead compound toxicity. Toxicity prediction algorithms frequently fragment molecules as they are based on the idea that the toxicity of a molecule is associated with particular structural motifs or groups within a compound—a defined “toxicophore” or chemotype [38]. A computational screening approach using compound fragment toxicophores was able to identify and classify mutagenic compounds with an error rate of 18 %—on the order of that of the experimental Ames test (with an error rate of 15 %). Fragmentation-based statistical analysis of toxicity databases has been performed to find substructure fragments, which are common and frequently found in toxic compounds and drugs. One group [39] used fragments from the Registry of Toxic Effects of Chemical Substances database (RTECS) as a training set. Use of a fragment-based toxicity predictor enables extension of experimental drug-like compound toxicity data since experimental data is only available for a small subset of compounds.

Fraggle, [40] the fragment store dataset is a database ([http://bioinf-applied.charite.de/fragment\\_store/](http://bioinf-applied.charite.de/fragment_store/)) that provides property information (charge, hydrophobicity, and binding site preferences) and performs statistical analysis and can view the IDS of drugs and toxic compounds, which contain the fragments. This website also has a fragment assembler. The database is based on the fragmentation of metabolic compounds, toxic compounds, and drugs, with

binding site preferences determined for each fragment using PDB (Protein Data Bank) structures. Often there is no shared overlap between the dataset of metabolic compounds compiled from KEGG [41] and the SuperDrug drugs [42] and SuperToxic database (<http://bioinformatics.charite.de/supertoxic/>) drugs, until they are fragmented and then there is a great deal of overlap of the fragments derived from these three databases [43].

There are useful techniques to assess toxicology and chemical risk assessment based on fragment substructural analysis. MULTICASE (<http://multicase.com/>) predicts carcinogenicity and toxicity on linear fragments, identification and use of substructure toxicophores, based on the recognition of chemical features and molecular descriptors. Substructures can predict specific compound activities, such as confer bitter taste, predict CNS activity, and predict plasma permeability.

Fragment design can be used in a negative way in that structural alerts and model building features in programs such as LeadsScope, Derek, and ToxTree [44] can be used to determine structural features and fragments to omit from fragment libraries. LeadsScope can be used to filter fragments according to nine global molecular properties (A log P; H bond acceptors and donors; Lipinski Score, MW, Parent atom count, polar surface area, rotatable bonds) and the compounds. LeadsScope contains 27,000 structural fingerprint features, which can be used for clustering and activity prediction of fragments, structure and similarity searching. The program has a recursive-partitioning algorithm to classify structural fragment subsets within a database and simulated annealing can be used to find statistical correlations within the subset based on the presence of sub structural features. LeadsScope can therefore be used to break down molecules into structural fragments and then use them for dynamic scaffold generation.

---

## 5 Conclusion: Successful Drugs from Fragments

Tipranavir (HIV protease inhibitor) reached the market in 2005 and was recognized as one of the first drugs developed by structure-based drug design starting from a fragment-based approach. In 2011, Vemurafenib (B-Raf inhibitor) reached the market [45]. Plexxxikon's Vemurafenib, a selective inhibitor of mutant kinase B-Raf (V600E) was the first drug developed by fragment screening and optimization. Beginning with a screen of diverse 20,000 fragment scaffold like compounds (MW 150–350 kDa) with activity assays performed at 200 μM (very high) concentrations, and with less than 8 H bond donors and acceptors. Seven azaindole fragments were found that inhibited three kinases PIM1, p38, and CSK by 30 %. The fragment was incorporated into a 3-aminophenyl analog and then inhibited PIM1 kinase in a single binding mode.

Computational methods in fragment-based drug design have proved to be successful in their application to numerous targets—kinases, proteases, dehydrogenases. Plexxikon (indeglitazar, PLX-4032), Abbott/AbbieVie (ABT 263, ABT518, ABT869, ABT737), Astex (AT9283, AT7519), Lilly/Protherics (LY517717), Vernalis/Novartis (VER52296/NVP-AUY-922), deCODE/Emerald (DG051), Locus (LP261) and SGX (SGX523), and Sunesis (SNS314) are all examples of drugs developed by fragment methods in clinic or development. Over the last 10 years fragment-based drug design has led to 50 small molecule hits that have advanced to lead structures [46].

In conclusion, computational fragment-based methods have demonstrated that they provide for greater diversity in compound hit development. Astek, Vernalis, Evotec, and Abbot all have had Hsp90 inhibitor development programs utilizing fragment-based drug design and although both Vernalis and Astek had initial phenol and resorcinol fragment hits, evolution of these fragment hits have taken very different paths even when similar initial fragment hits are identified indicating the value of fragment-based drug design in providing for compound diversity [47]. More novel computational fragment-based methods will facilitate further drug design in the future.

## References

- Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274:1531–1534
- Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent developments in fragment-based drug discovery. *J Med Chem* 51: 3661–3680
- Mazanetz M, Law R, Whittaker M (2014) Hit and lead Identification from fragments. In: Scheider G (ed) *De novo molecular design*. Wiley, New York, pp 143–200
- Davis BJ, Erlandson DA (2013) Learning from our mistakes: the ‘unknown knowns’ in fragment screening. *Bioorg Med Chem Lett* 23:2844–2852
- Murray CW, Verdonk ML, Rees DC (2012) Experiences in fragment-based drug discovery. *Trends Pharmacol Sci* 33:224–232
- van Deursen R, Blum LC, Reymond JL (2011) Visualization of the chemical space of fragments, lead-like and drug-like molecules in PubChem. *J Comput Aided Mol Des* 25: 649–662
- Blum JC, van Deursen R, Reymond JL (2011) Visualisation and subsets of the chemical universe dataset GDB-13 for virtual screening. *J Comput Aided Mol Des* 25:637–647
- Polishchuk PG, Madzhidov TI, Varnek A (2013) Estimation of the size of drug-like chemical space based on GDB-17 data. *J Comput Aided Mol Des* 27(8):675–679
- Bohm HJ (1992) LUDI: rule based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des* 6:593–606
- Matto C, Bellamacina CR, Peisach E, Pereira A, Vitkup D, Petsko GA, Ringe D (2006) Multiple solvent crystal structures: probing binding sites, plasticity and hydration. *J Mol Biol* 357:1471–1482
- Raman EP, Yu W, Guvench O, MacKerell AD Jr (2011) Reproducing crystal binding modes of ligand functional groups using site-identification by ligand competitive saturation (SILCS) simulations. *J Chem Inf Model* 51:877–896
- Ngan CH, Bohnuud T, Mottarella SE, Beglov D, Villar EA, Hall DR, Kozakov D, Vajda S (2012) FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res* 40:W271–W275
- Lewell XQ, Judd DB, Watson SP, Hann MM (1998) RECAP—retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments

- with useful applications in combinatorial chemistry. *J Chem Inf Comput Sci* 38:511–522
14. Cooper T, Andrews-Cramer K (2000) Designed chemical libraries for hit/lead optimization. *Innovat Pharm Technol* 5:46–53
  15. Cramer RD, Soltanshahi F, Jilek R, Campbell B (2007) AllChem: generating and searching 10(20) synthetically accessible structures. *J Comput Aided Mol Des* 21:341–350
  16. Rarey M, Dixon JS (1998) Feature trees: a new molecular similarity measure based on tree matching. *J Comput Aided Mol Des* 12: 471–490
  17. Lounkine E, Auer J, Bajorath J (2008) Formal concept analysis for the identification of molecular fragment combinations specific for active and highly potent compounds. *J Med Chem* 51:5342–5348
  18. Batista J, Godden JW, Bajorath J (2006) Assessment of molecular similarity from the analysis of randomly generated structural fragment populations. *J Chem Inf Model* 46: 1937–1944
  19. Khashan R (2012) FraVLib a free database mining software for generating “Fragment-based Virtual Library” using pocket similarity search of ligand-receptor complexes. *J Cheminform* 4:18
  20. de Graaf C, Vischer HF, de Kloet GE et al (2013) Small and colorful stones make beautiful mosaics: fragment-based chemogenomics. *Drug Discov Today* 18(7–8):323–330
  21. Yuling A, Sherman W, Dixon SL (2012) Hole filling and library optimization: application to commercially available fragment libraries. *Bioorg Med Chem* 20:5379–5387
  22. Sirci F, Enade P, Istyastono HF et al (2012) Virtual fragment screening: discovery of histamine H3 receptor ligands using ligand-based and protein-based molecular fingerprints. *J Chem Inf Model* 52:3308–3324
  23. Favia AD, Bottegoni G, Nobeli I et al (2011) SERAPhiC: a benchmark for in silico fragment-based drug design. *J Chem Inf Model* 51:2882–2896
  24. Zauhar RJ, Gianti E, Welsh WJ (2013) Fragment-based shape signatures: a new tool for virtual screening and drug discovery. *J Comput Aided Mol Des* 27:1009–1036
  25. Hoffer L, Renaud JP, Horvath DJ (2013) In silico fragment-based drug discovery: setup and validation of a fragment-to-lead computational protocol using S4MPLE. *J Chem Inf Model* 53(4):836–851
  26. Gillet V, Johnson AP, Mata P et al (1993) Sprout: a program for structure generation. *J Comput Aided Mol Des* 7:127–153
  27. Day F, Caflisch A (2008) Fragment based de novo ligand design by multiobjective evolutionary optimization. *J Chem Inf Model* 2008(48):679–690
  28. Yuan H, Tai W, Hu S et al (2013) Fragment-based strategy for structural optimization in combination with 3D-QSAR. *J Comput Aided Mol Des* 27:897–915
  29. Durrant JD, Lindert S, McCammon JA (2013) Autogrow 3.0: an improved algorithm for chemically tractable, semi-automated protein inhibitor design. *J Mol Graph Model* 44: 104–112
  30. Durrant JD, McCammon JA (2012) AutoClickChem: click chemistry in silico. *PLoS Comput Biol* 8:e1002397
  31. Pierce AC, RaoG BGW (2004) BREED: generating novel inhibitors through hybridization of known ligands, applications to CDK2, p38 and HIV protease. *J Med Chem* 47:2768–2775
  32. Lindert S, Durrant JD, McCammon JA (2012) LigMerge: a fast algorithm to generate models of novel potential ligands from sets of known binders. *Chem Biol Drug Des* 80:358–365
  33. Chung S, Parker JB, Bianchet M, Amzel LM, Stivers JT (2009) Impact of linker strain and flexibility in the design of a fragment-based inhibitor. *Nat Chem Biol* 5:407–413
  34. Thompson DC, Denny RA, Nilakantan R et al (2008) CONFIRM: connecting fragments found in receptor molecules. *J Comput Aided Mol Des* 22:761–772
  35. Maass P, Schulz-Gasch T, Stahl M, Rarey M (2007) Recore: a fast and versatile method for scaffold hopping based on small molecule crystal structure conformations. *J Chem Inf Model* 47:390–399
  36. Vainio MJ, Kogej T, Raubacher F, Sadowski J (2013) Scaffold hopping by fragment replacement. *J Chem Inf Model* 53:1825–1835
  37. Kump JL III, Blumenthal SN, Wang Q et al (2012) A fragment-based approach to the SAMPL3 Challenge. *J Comput Aided Mol Des* 26:583–594
  38. Kazius J, McGuire R, Bursi R R (2005) Derivation and validation of toxicophores for mutagenicity prediction. *J Med Chem* 48: 312–320
  39. von Korff M, Sander T T (2006) Toxicity-indicating structural patterns. *J Chem Inf Model* 46:536–544
  40. Ahmed J, Worth CL, Thaben P et al (2011) Fragment Store—a comprehensive database of fragments linking metabolites, toxic molecules and drugs. *Nucleic Acids Res* 39: D1049–D1054

41. Kaneshisa M (2002) The KEGG database. Novartis Found Symp 247:91–101
42. Wisharte DS, Knos C, Guo AC et al (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 36:D901–D906
43. Schmidt U, Struck S, Gruening B et al (2009) SuperToxic: a comprehensive database of toxic compounds. Nucleic Acids Res 37:D295–D299
44. Roberts G, Myatt GJ, Johnson WP et al (2000) Leadscape: software for exploring large sets of screening data. J Chem Inf Comput Sci 40: 1302–1314
45. Erlanson DA (2012) Introduction to fragment based drug discovery. Top Curr Chem 317:1–32
46. Bollag G, Tsai J, Zhang J et al (2012) Vemurafenib: the first drug approved for BRAF-mutant cancer. Nat Rev Drug Discov 11:873–886
47. Rougley SD, Hubbard RE (2011) More novel computational fragment based methods will facilitate further drug design in the future. J Med Chem 54:3989–4005

# Chapter 11

## Design Strategies for Computational Fragment-Based Drug Design

Zenon D. Konteatis

### Abstract

The computational design method described in this chapter is an approach to de-risking the design process due to the limitations of current computational algorithms with respect to predictive accuracy. The method takes advantage of the crystallographically demonstrated interactions between a ligand and its protein target, and through systematic, one fragment replacements allows for quick feedback on the direction of the designs. This design approach can still be useful in the future as computational algorithms improve and become more predictive and reliable.

**Key words** Computational fragment-based design, Design strategies, Molecules from fragments, Fragment designs, Fragment evolution, Chemotype evolution, Molecular evolution

---

### 1 Introduction

Structure-based drug design (SBDD) approaches have become well established in drug discovery over the last 25 years [1–4] and have contributed to the successful introduction of more than 40 New Chemical Entities (NCEs) in clinical trials [5] and eleven marketed drugs [6]. Most pharmaceutical and biotechnology companies have SBDD departments incorporating X-ray crystallography, NMR, and computational groups, and recently many academic laboratories have developed this discovery capability [7].

Experimental fragment-based drug design (experimental FBDD) emerged in the late 1990s as a new approach to drug discovery [8], and it has gained wide acceptance over the last fifteen years as multiple pharmaceutical and biotechnology companies and academic institutions have incorporated experimental FBDD in their research capabilities [9]. Multiple NCE's developed by the use of experimental FBDD have reached clinical trials [9] and one of them was approved by the FDA in 2011 [10]. Experimental FBDD requires specialized methods such as NMR, X-ray crystallography, mass spectroscopy, surface plasmon resonance (SPR),

tethering, isothermal calorimetry (ITC), and high concentration assay (HCA) screening to detect the weak affinity of fragments for their protein targets [11].

Computational technologies developed over the last three decades have contributed to the success of SBDD and have been recently adapted to establish computational fragment-based drug design (computational FBDD) [12, 13]. Computational FBDD methods use the same principles established for experimental FBDD and are beginning to aid in the successful design of FBDD-based inhibitors for a variety of pharmaceutical targets [13]. In this chapter we discuss various strategies employed in designing ligands for specific protein targets starting with a co-crystal structure, irrespective of the specific computational software used.

---

## 2 Materials: Strengths and Limitations of Computational Software Programs for Drug Design

A variety of available computational software programs that enable docking or energy minimization of a ligand in the protein binding site (e.g., CCG's MOE, Schrödinger's Glide, GOLD, FLEXX) can be used to prosecute the design approaches described in this chapter. A crystal structure of a ligand with a known binding affinity for the protein target is also necessary to enable these design strategies.

### 2.1 Binding Pose Accuracy

A number of computational methods have been used in computational SBDD and FBDD with varying degrees of success in selecting or designing drug-like compounds. A variety of investigations on the accuracy of these methods [14–17] have led to the conclusion that current software generate ligand conformations and binding poses similar (<2 Å rmsd) to the ones observed in co-crystal structures. Predicting the correct binding mode of ligands and fragments is of great value in FBDD, especially when combined with experimental FBDD to focus the number of fragments or ligands to be examined by a biophysical method such as NMR or crystallography [13].

### 2.2 Affinity Prediction Limitation

The same studies also demonstrated that docking programs with scoring functions do not predict ligand binding affinities or rank order ligands correctly [14–17]. When applied to a congeneric series of compounds interaction energies showed a modest correlation ( $r^2$  of 0.7) with experimental affinities [18], establishing the upper limit of prediction for current algorithms. This accuracy limitation can be attributed to the absence of a robust solvation treatment and the lack of full accounting for the entropy effects of both the ligand and the protein. Several methods with more rigorous physical treatment of the thermodynamic cycle, such as free-energy perturbation (FEP) [19], thermodynamic integration (TI) [20],

and grand canonical Monte Carlo (GCMC) [21] simulations have the potential to generate more quantitatively accurate rank ordering of ligands. However, to date not enough data have been reported in the literature to make a proper assessment. This inaccuracy in rank ordering predicted affinities of virtual compounds hampers the practical application of computational FBDD and SBDD and has limited virtual screening campaigns in SBDD to just enrichment of active compounds over inactive ones [22, 23].

### 2.3 Multiple Binding Modes for Fragments

One of the challenges that computational fragment-based design strategies face is the prediction of binding modes [21, 24]. In general the energy difference between two binding modes for a fragment are much smaller than that seen for full molecules and this gives rise to multiple binding modes for fragments [25]. However, predicting not only the lowest energy pose but other reasonable poses can be a true strength in this strategy as these other poses may be better suited for building full molecules by joining or linking with nearby poses of other fragments [13].

---

## 3 Methods: General Design Strategies

A large number of computational design methods have been described in the literature with varying degrees of success. The methods discussed herein have proven to decrease the risk of design failure to an extent that overcomes the limitations mentioned in Subheading 2.

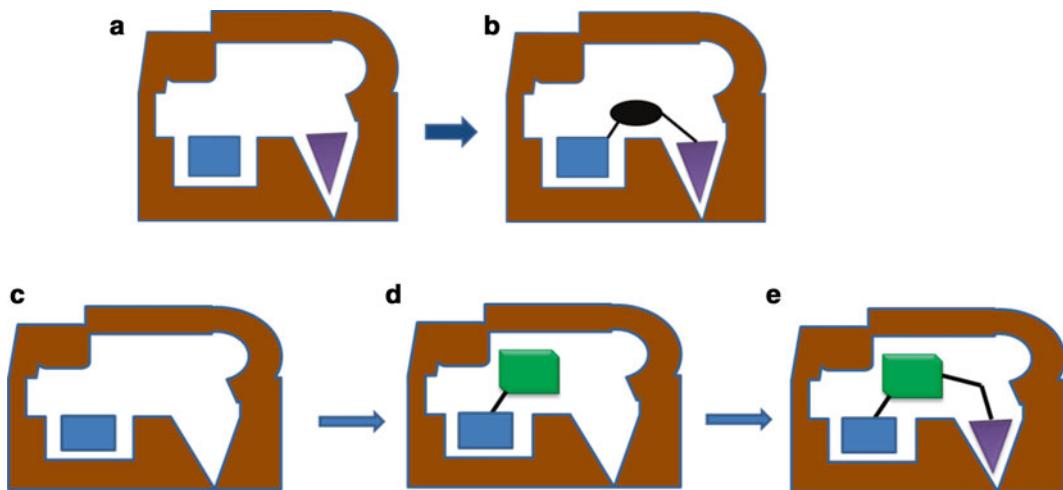
### 3.1 FBDD Similarities and Builds

Computational FBDD is based on many of the same concepts that experimental FBDD has used successfully: starting with small chemical building blocks (fragments) both methods try to build novel compounds with “drug-like” characteristics by growing, directly joining, or linking fragments as shown in Fig. 1.

Computational FBDD methods that can dock a number of fragments independently in the binding pocket of the target protein (receptor) can easily implement both linking and growing approaches. Other methods that only dock one fragment at a time can also be used by first docking a key fragment and then evolving this fragment by building extensions or joining a second fragment and re-docking the combined two fragment molecule and so on until the full molecule is elaborated.

### 3.2 Knowledge-Based (Pharmacophore) Screening

Early de novo design and Virtual Screening approaches relied exclusively on computation and the inaccuracy of scoring functions led to limited success as already discussed. In this section we will show that incorporating experimental data in computational fragment-based design approaches can both complement and de-risk the process.



**Fig. 1** Fragment-based drug design: linking, growing. Linking: (a) Fragment 1 binds to the target protein at site 1; fragment 2 binds at an adjacent site 2. (b) Fragments 1 and 2 are joined together by a linking group that generates a molecule which spans both sites (adapted from ref. 26). Growing: (c) Fragment 1 binds to the target protein at site 1; (d) a second fragment is joined directly to fragment 1; (e) and the molecule is completed by growing into site 2

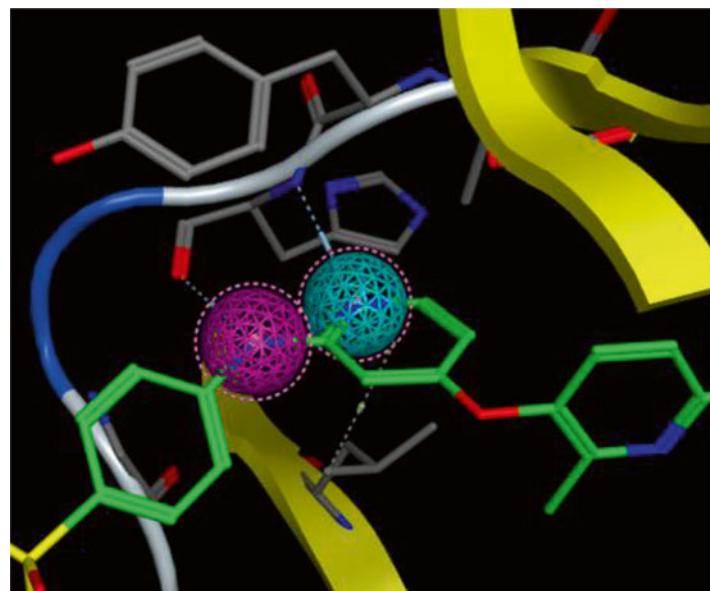
A number of studies have demonstrated that virtual screening can provide successful lead molecules when combined with a pharmacophore-directed search, especially when the pharmacophore is based on a co-crystal structure of a ligand with its receptor. Many examples exist in the area of kinase inhibition [13, 15, 18], and one such example is illustrated in Fig. 2, with TGF-beta kinase. A crystallographic starting point (PDB code: 2WOU) is used to generate a pharmacophore query aimed at the protein-ligand hinge binding interactions and the virtual screening identifies molecules that contain such binding motifs and fit the protein binding pocket at least geometrically.

### 3.3 Computational Fragment-Based Design Evolution (Chemotype Evolution)

Computational approaches can be made more successful by adding knowledge-based experimental information to the design process. The pharmacophore example explained above demonstrates that most computational algorithms, which are very good at producing meaningful poses [14–17], and if combined with experimental data can generate designs that are very productive. We term this process Computational Fragment-Based Design Evolution (cFBDE). This is composed of four key steps that take the designs from a known ligand to novel compounds.

#### 3.3.1 Co-crystal Structure as a Normalizing Factor and Relative (Not Absolute) Designs

At the center of this method is the availability of experimental data for the system of study, especially a co-crystal structure of an active molecule in the binding pocket of the protein target. We use the co-crystal structure as a *normalizing factor* to de-risk the designs



**Fig. 2** Pharmacophore selection for TGF-beta. Hydrogen bond acceptor (cyan) and donor (magenta) pharmacophoric elements are generated to direct the virtual screening to identify molecules that will bind to the hinge region of TGF-beta kinase (PDB code 2WOU). Figure generated in MOE [27]

and thus produce *relative* designs *not absolute* designs. By using the known inhibitor and its interactions with the protein, we start with *real interactions* which then can be the *reference point* for computational comparisons with other computationally driven interactions. Thus, computationally reproducing the co-crystal ligand binding pose becomes the first step in this process.

### 3.3.2 Stepwise Fragment Replacement Strategy to Address Computational Limitations

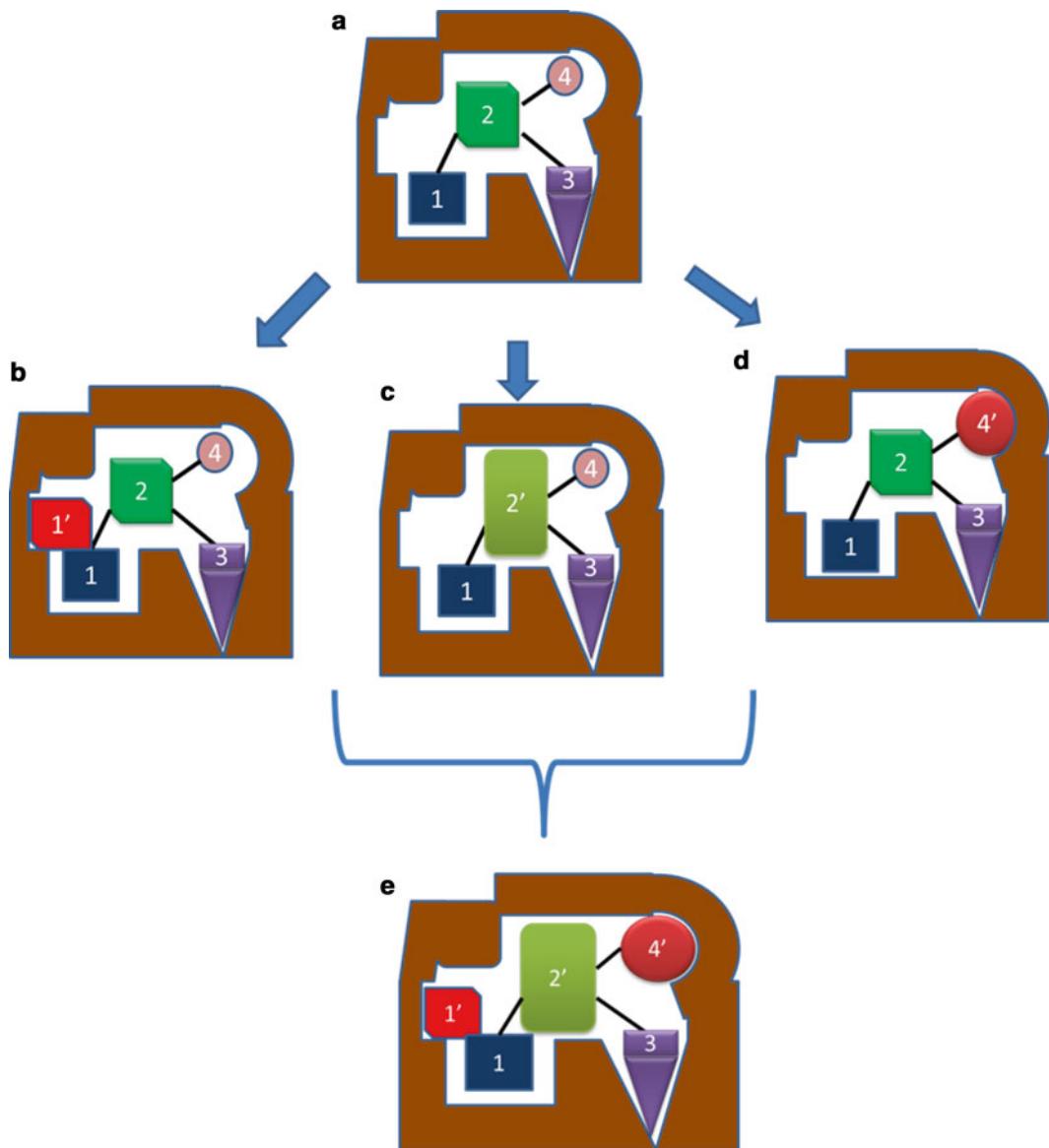
The co-crystal structure ligand is considered to be composed of a number of fragments. One of these fragments is replaced by another fragment while the rest of the ligand remains the same; this enables fragment optimization without changing the linking scheme, thus avoiding one of the most difficult steps in fragment-based design. Each fragment in the molecule is replaced independently to generate multiple new subclasses of the original ligand, as illustrated in Fig. 3a–d. Multiple fragment replacements can be tried, and the best are carried forward.

### 3.3.3 Pharmacophore-Driven New Interactions (Extensions)

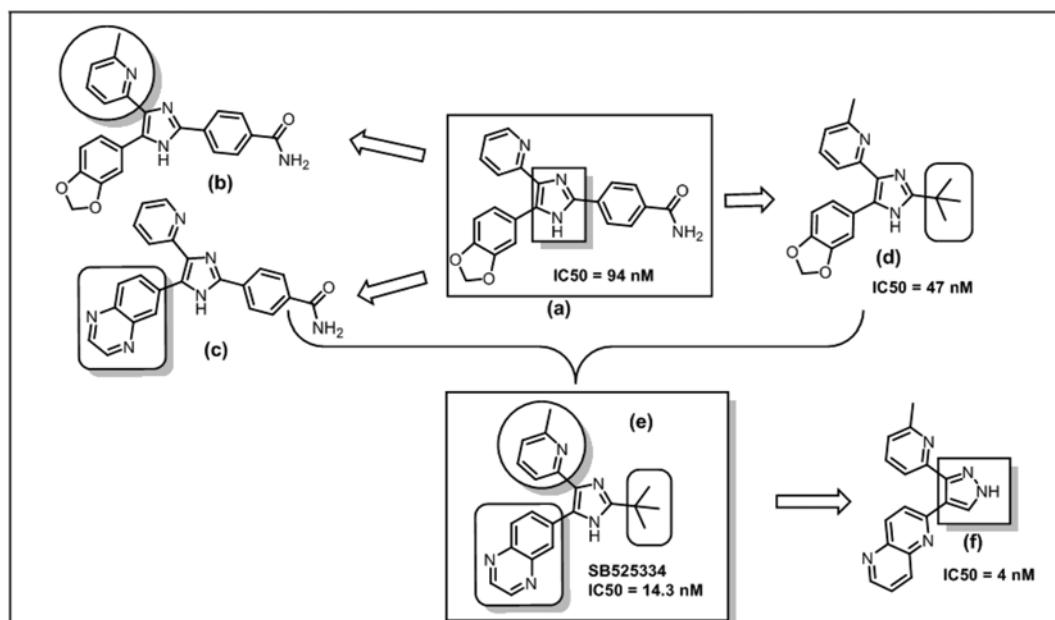
During this fragment replacement, new interactions with the protein may be targeted with the new fragment, as seen in Fig. 3b, thus creating new, pharmacophore-driven extensions that have a very good chance of adding affinity to the molecule produced.

### 3.3.4 Design Recombination to De Novo Designs

Once multiple optimized fragments are produced for each position, they can be recombined to generate completely new classes of compounds as depicted in Fig. 3e. These molecules are novel



**Fig. 3** Stepwise fragment replacement generates new subclasses of ligands. Recombination leads to new chemotypes equivalent to de novo designs. (a) Crystallographic ligand composed of four fragments is first reproduced computationally in the protein binding pocket by docking or other algorithm in use; (b) fragment 1 is extended 1' to add a new, nearby interaction with the protein (fragment optimization or extension); (c) fragment 2 is replaced by another fragment 2' (scaffold hopping); (d) fragment 4 is optimized to fragment 4'; (e) recombining all the new, successful individual replacements produces a new, novel class or chemotype (de novo design)



**Fig. 4** Design process generates potent TGF-beta inhibitors. (a) A crystallographic ligand with IC<sub>50</sub> of 94 nM was redesigned fragment (b) by fragment (c) and (d) to produce a recombined inhibitor (e) with 14.3 nM affinity for TGF-beta. Scaffold hopping by changing the imidazole fragment to a pyrazole (f) and re-optimizing the hinge binding fragment (c) gave a new compound class with 4 nM affinity

classes that are the equivalent of de novo produced designs but with much less computational risk.

The cFBDE process is illustrated above in Fig. 4 with an example from the TGF-beta reported literature. Each replaced fragment is highlighted and the stepwise process depicted by arrows.

#### 4 Conclusion

The overall design strategy described here is a reliable and efficient method of generating new, novel chemotypes starting with a known ligand and its co-crystal binding site. The method can be also used in cases where a co-crystal structure is not available but a reasonable receptor homology model can be built using experimental information of related protein class members and a few ligands with known structure-activity relationship (SAR) in the compound class of interest. In the case where an apo crystal structure is available, but not a co-crystal with a ligand, then careful docking experiments can be used to set up a reasonable model of the ligand binding site in the receptor by analogy to similar receptors with known ligand binding sites. Thus, this methodology can be applied in many cases involving most receptor classes where members of the class have produced some co-crystal structures.

This design strategy can be further de-risked by obtaining co-crystal structures at key intervals to guide redesigns based on protein adaptation to related states that may influence ligand binding and affinity.

## References

1. Roberts N, Martin J, Kinchington D et al (1990) Rational design of peptide-based HIV proteinase inhibitors. *Science* 248:358–361
2. Erickson J, Neidhart D, VanDrie J et al (1990) Design, activity and 2.8 Å crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease. *Science* 249:527–533
3. Hartman GD, Wgbertson MS, Halczenko W et al (1992) Non-peptide fibrinogen receptor antagonists. I. Discovery and design of exosite inhibitors. *J Med Chem* 35:4649–4652
4. Dorsey BD, Levin RB, McDaniel SL et al (1994) L-735,524: the design of a potent and orally available HIV inhibitor. *J Med Chem* 37:3443–3451
5. Hardy LW, Malikayil A (2003) The impact of structure-guided drug design on clinical agents. *Curr Drug Discov* 3:15–20
6. Perola E (2010) An analysis of the binding efficiencies of drugs and their leads in successful drug discovery programs. *J Med Chem* 53: 2986–2997
7. Anderson AC (2003) The process of structure-based drug design. *Chem Biol* 10:787–797
8. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274: 1531–1534
9. Murray CW, Verdonk ML, Rees DC (2012) Trends Pharmacol Sci 33:224–232
10. Bollag G et al (2010) Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* 467:596–599
11. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) *J Med Chem* 51:3662–3680
12. Konteatis Z (2010) *In silico* fragment-based drug design. *Expert Opin Drug Discov* 5:1047–1065
13. Loving K, Alberts I, Sherman W (2010) Computational approaches for fragment-based and *de novo* design. *Curr Top Med Chem* 10:14–32
14. Warren GL, Andrews CW, Capelli A-M et al (2006) A critical assessment of docking programs and scoring functions. *J Med Chem* 49:5912–5931
15. Lyne PD, Lamb ML, Saeh JC (2006) Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J Med Chem* 49:4805–4808
16. Sandor M, Kiss R, Keseru GM (2010) Virtual fragment docking by Glide: a validation study on 190 protein-fragment complexes. *J Chem Inf Model* 50:1165–1172
17. Chen Y, Shoichet BK (2009) Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat Chem Biol* 5:358–364
18. McDaughay GM, Culberson JC, Feuston BP et al (2006) Scoring of KDR kinase inhibitors: using interaction energy as a guide for ranking. *Mol Divers* 10:341–347
19. Kollman PA (1993) Free energy calculations: applications to chemical and biochemical phenomena. *Chem Rev* 93:2395–2417
20. van Gunsteren WF, Berendsen HJC (1987) Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry. *J Comput Aided Mol Des* 1:171–176
21. Clark M, Guarneri F, Shkurko I, Wiseman J (2006) Grand canonical Monte Carlo simulation of ligand-protein binding. *J Chem Inf Model* 46:231–242
22. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* 47:1750–1759
23. Klon AE, Glick M, Davies JW (2004) Combination of a naïve Bayes classifier with consensus scoring improves enrichment of high-throughput docking results. *J Med Chem* 47:4356–4359
24. Marcou G, Rognan D (2007) Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J Chem Inf Model* 47:195–207
25. Jahnke W, Erlanson DA (2006) Fragment-based approaches in drug discovery. Wiley-VCH, Weinheim
26. Rees DC, Congreve M, Murray CW, Carr R (2004) Fragment-based lead discovery. *Nat Rev Drug Discov* 3:660–672
27. Chemical Computing Group Inc. (2013) Molecular operating environment (MOE). Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7

# Chapter 12

## Protein Binding Site Analysis for Drug Discovery Using a Computational Fragment-Based Method

Jennifer L. Ludington

### Abstract

One of the most powerful tools for designing drug molecules is an understanding of the target protein's binding site. Identifying key amino acids and understanding the electronic, steric, and solvation properties of the site enables the design of potent ligands. Of equal importance for the success of a drug discovery program is the evaluation of binding site druggability. Determining, *a priori*, if a particular binding site has the appropriate character to bind drug-like ligands saves research time and money.

While there are a variety of experimental and computational techniques to identify and characterize binding sites, the focus of this chapter is on Binding Site Analysis (BSA) using virtual fragment simulations. The methodology of the technique is described, along with examples of successful application to drug discovery programs. BSA both indicates if a protein is a viable target for drug discovery and provides a roadmap for designing ligands. Using a computational fragment-based method is an effective means of understanding of a binding site.

**Key words** Protein binding site, Binding site analysis, Fragment-based drug design, Structure-based drug design, Binding site identification, Binding site characterization, Druggability

---

### 1 Introduction

Understanding protein binding sites is the basis of structure-based drug design. A binding site is a region on a protein where another molecule interacts. A binding site may be the interface of a protein–protein interaction, a region where a ligand co-factor binds, or the catalytic pocket of an enzyme in which a molecule is cleaved or modified. A drug molecule is intended to affect the function of a protein *in vivo*, so it must either bind in a region overlapping with that of the natural ligand (to either mimic or inhibit that ligand), or in an allosteric pocket that affects the shape of the natural ligand's binding site. A protein is considered to be druggable as defined by Hopkins and Groom if a ligand with “drug-like” properties (i.e., orally bioavailable) can bind with high affinity and if modification of the protein's function positively affects a disease state [1]. The presence of a druggable binding site is a determining

factor for whether or not a protein is a viable target for drug discovery. The knowledge that a protein lacks a druggable binding site early in a drug discovery project frees a research team to pursue alternative targets.

---

## 2 Methodologies for Binding Site Analysis

### 2.1 Techniques for Binding Site Identification

Binding Site Analysis (BSA) is a two stage process, consisting of an identification stage in which potential ligand binding sites are located and a characterization stage in which the identified sites are evaluated for druggability. Various experimental and computational techniques have been developed for binding site identification and druggability determination as discussed extensively in recent reviews [2–5]. Some techniques are only applicable to the identification stage, while other methods like fragment-based approaches, give insight into both stages of BSA.

The most definitive method of determining where a ligand binds is by experimentally determining the 3D structure of a protein–ligand complex, but this approach requires a ligand for each binding site to be located. Three general approaches to computational binding site identification that do not rely on knowledge of the ligand are geometric, energy-based, and fragment-based methods. A 2010 study by Schmidtke et al. [6] evaluated the performance of four algorithms with a large-scale data set; the geometric algorithms SiteFinder [7] and fpocket [8] and the energy-based methods SiteMap [9, 10] and PocketFinder [11] were compared. The techniques performed similarly when evaluating the 5 highest-scored pockets for *holo* structures; all had a 95 % or higher success rate of identifying a true binding site [6]. The success rate of the methods for identifying the known site as the first ranked pocket was between 70 and 82 % for *holo* structures and between 42 and 80 % for *apo* structures [6]. It appears to be more difficult to identify the binding sites of *apo* structures and yet this is where the computational methods are needed the most.

The computational fragment-based approach to BSA used in the Locus Pharmaceuticals Core Technology (LCT) [12, 13] is analogous to the experimental methodology, Multiple Solvent Crystal Structures (MSCS), developed by Ringe and coworkers [14–16]. With the MSCS method, X-ray structures are solved for protein crystals after they are equilibrated in an organic solvent. Superimposing the crystal structures from various solvent soaks overlays the binding locations of the different solvent molecules. The region where many types of solvent molecules cluster, irrespective of size and polarity, is indicative of a binding site. Ringe also reported three categories of water molecules: tightly bound or structural waters that were present in all structures, partially ordered waters that were only present in one of the structures, and fully disordered waters that could not be

**Table 1**  
**Example probe fragments used for binding site identification**

Acetamide	Dimethylsulfoxide	Methanol
Acetone	Ethane	Methylamine
Benzene	Ether	Pyrimidine
Carboxylic acid	Imidazole	Urea

resolved in any of the crystal structures [14]. The waters observed in the sites where different organic solvent molecules overlapped were of the second type, displaceable and unlike fully disordered water capable of gaining entropy when displaced [14].

The LCT-BSA methodology is based on Grand Canonical Monte Carlo (GCMC) simulations of fragments with a protein, as presented in detail in previous publications [17–19]. Similar to MSCS, the simulations begin with saturating the protein simulation cell with a solvent of an organic fragment [17–19]. Both the fragment and the protein are treated as rigid bodies, so multiple conformations of flexible fragments are simulated individually [17–19]. As the simulation progresses, a unitless parameter B (related to the excess chemical potential in the system) is lowered, causing fragments to leave the protein simulation box until only the tightest-binding fragments remain [17–19]. A diverse set of probe fragments including water is simulated, and the results of the separate simulations are combined [2, 20, 21]. These fragments are small to prevent steric exclusion from legitimate pockets [18]. The probe fragments represent the different types of interactions a ligand can have with a protein: hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), aliphatic hydrophobic, and aromatic [21]. Examples of probe fragments for identifying binding sites are shown in Table 1. Putative binding sites are located where the different fragment types cluster with high binding affinity, identifying a hotspot [2, 20, 21]. (The term hotspot was coined to refer to amino acids whose interactions provide most of the binding affinity in protein–protein interactions [22]. The term is generalized here for protein–ligand interactions.) For a hotspot to represent a binding site, the waters in this region must be less tightly bound to the protein than the other fragments are [20, 21]; they should be analogous to the type II waters of MSCS [14].

The accuracy of the GCMC calculations comes at a computational expense that makes it unrealistic to test with the large data sets used to validate faster techniques. However, this approach has been successful for the limited set of test cases used. Anecdotal evidence from drug discovery programs has also demonstrated success. Data have been published on the identification of binding sites for dihydrofolate reductase (DFHR) and

human immunodeficiency virus aspartic protease (HIV protease) [2]. A calculated hotspot that was identified on DHFR by a computational fragment cluster was found to be located in the same region of the protein where an electron is transferred from nicotinamide adenine dihydrogen phosphate (NADPH) to 7, 8-dihydrofolate (DHF) [2]. In the case of HIV protease, fragment probes simulated with a mutant form of the protein predicted that the extended dimer interface is a significant binding region [2]. The result was consistent with the overlay of the experimental binding positions of six HIV protease inhibitors [2]. Binding locations and binding free energies were calculated for T4 lysozyme [18], which has a tight, induced pocket for small organic molecules. All but 3 of the 14 test molecules were found in the binding site and the calculated lowest energy pose for each found molecule was in good agreement with the experimental location, where available [18]. The rank-order of molecules based on the calculated free energies of binding also matched well with the experimental results [18]. These test cases support the assumption that high accuracy binding calculations give high success rates.

Protein flexibility affects binding site identification as demonstrated by the better success rates using structures with bound ligands compared to *apo* structures. A tight pocket can exclude fragments in a significant region, as was seen in the study of T4 lysozyme [18]. Protein–protein interaction (PPI) sites have been recognized to be more flexible than protein–ligand sites, as reviewed by Grimme et al. [23]. Therefore, the protein conformation used in the simulations of a PPI site is critical. A conformation omitting a transient, deeper pocket would greatly influence the BSA and ligand design. Small-motion flexibility is addressed indirectly with LCT; simulations generate ensembles of fragments instead of single-point binding modes. Also, the binding site is defined as a consensus site where multiple fragments bind with high affinity, so the exclusion of a few of the fragment types will not typically prevent binding site identification. Molecular dynamics (MD) and normal mode analysis (NMA) can be included as part of the LCT protein preparation to address larger scale changes in protein conformation [21, 24].

## 2.2 Druggability of Protein Binding Sites

Determining the druggability of a protein binding site is a challenging aspect of BSA that has been addressed with a number of approaches. Protein druggability is not always defined consistently and non-druggable proteins can be recategorized if new ligands are found [5]. Screening hit rates from high-throughput [25], NMR-fragment [26], or computational-fragment [27] methods have been used successfully as measures of druggability. Several methods have been developed that predict druggability based on calculated physicochemical descriptors of the protein pockets [10, 26, 28–31]. These methods tend to rate larger pocket size, solvent shielding, and hydrophobicity as factors which increase

druggability, but differ in whether polar and hydrophilic contact areas are considered favorable [5]. The LCT method uses binding information for sets of fragments with homogenous chemical character to evaluate the druggability of a binding site [20, 21].

---

### 3 Fragment-Based Binding Site Analysis

#### 3.1 Identification of Binding Sites

The identification phase of the Locus BSA process consists of finding clusters of fragments with high binding affinity and then determining which ones define bio-relevant hotspots. Hotspots are anchoring regions for ligand binding, and are a necessary component of a binding site [2, 20–22]. The ATP-site hinge region present in kinases is an example of such a hotspot, and fragment-based BSA identified this region in p38 MAP kinase [13].

The fragment clusters that represent an actual ligand binding site have the following attributes: they contain the majority of probe fragment types, they contain fragments with low energy binding poses, the fragments in the cluster can outcompete water binding, and they are located on regions of the protein where the topography allows for more than minimal surface contact [20, 21]. A hotspot will not represent a ligand binding site if tightly bound water is located in the same region, since a ligand cannot bind with high affinity if its interactions with the protein are not stronger than those of the water that will be replaced [21, 32]. The binding pocket should ideally contain some deep clefts so that the ligand can make adequate surface contact for tight binding [20, 21]. Explicit solvation is not part of these fragment simulations, so the solvent screening of solvent exposed interactions is not accounted for which may cause overestimated binding affinities of fragments. Consequently, sites that are significantly solvent exposed are excluded, since they are most likely artifacts of the technique. Fragment clusters in true binding sites are also persistent. The presence of these clusters is largely independent of the values of parameters such as the number of fragment types in a cluster and the cluster radius [21]. A binding site must also be solvent accessible so that ligands can enter the site. If an interesting fragment cluster is totally encapsulated by the protein, MD and NMA simulations can be used to explore if there is a state with reasonable energy in which the site is accessible [21, 24].

In addition to detecting the binding site of the natural ligand, BSA can identify novel sub-pockets and alternative sites. These types of regions allow for innovation in designed molecules, which is important for intellectual property and the exploration of ligand property space. In addition to the ATP and allosteric site, BSA of multiple p38 MAP kinase X-ray crystal structures identified three novel binding sites (the docking site, the C-lobe exosite, and the ACT site) [13]. Each of these sites was later confirmed experimentally [13].

### 3.2 Characterization of Binding Sites

The purpose of the BSA characterization phase is to determine the druggability of binding sites found in the identification stage. For a binding site to be druggable there must be ligands with “drug-like” properties that bind with high affinity and that modify the protein’s function [1]. To accommodate a ligand with acceptable properties, a binding site should be relatively compact, have some deeper clefts, and support strong interactions of varying chemical character [20, 21]. For PPI sites, finding this desired topography is challenging, as these sites may have a large, shallow surface area comprised of multiple weaker interactions [23]. If the majority of binding interactions are weak, it is difficult to design molecules with the necessary potency and still constrain ligand size. Although many PPI sites may not be druggable, a BSA can predict the exceptions.

To evaluate binding site druggability, fragments that belong to sets with homogeneous chemical character are simulated with the protein. Each characterization set contains fragments representing one type of interaction in the probe set (HBD, HBA, aliphatic hydrophobic, aromatic); additionally a set of fragments with multiple interaction types is simulated [20, 21]. The fragments in the characterization sets can be slightly larger than the probe fragments [21]. The fragment clusters for this phase of analysis each contain a single category of fragments, but several types of fragments in that category must bind with good affinity and not be excluded by tightly bound water [21]. If there is a water molecule with high affinity that outcompetes fragments in the binding pocket, a new protein may be simulated that includes that water as part of the structure [21, 32]. In that case, ligands can be designed to interact with the water molecule, so competition with the water is no longer an issue [21, 32].

To judge the druggability of a binding site, the following criteria are assessed. Several high affinity characterization clusters should be present within 15 Å of the hotspot fragment cluster [21]. If the site has a larger radius, it may indicate that only ligands with high molecular weights can bind with high potency. The fragment clusters should be in linkable distance and clusters of different chemical character should alternate throughout the site [21]. The ratio and pattern of hydrophobic and hydrophilic surface areas in the binding site affects the membrane permeability and solubility of designed ligands. A site for which complementary ligands will be linear or detergent-like will have poor druggability. Finally, the binding site should not be highly hydrated, according to the water fragment simulation [21]. These principles for evaluating the druggability of a binding site are not absolute, but are guidelines for finding the sites best suited for binding a drug-like molecule [21].

In addition to evaluating the druggability of a binding site, binding site characterization can generate strategies for ligand design. Analysis of the simulation data highlights promising molecular scaffolds and reveals which fragment types are acceptable in

each sub-pocket of the binding site. The characterization analysis of the allosteric binding site of p38 MAP kinase provided important design information. Previously, ligands that bound to the allosteric site of p38 had hydrophobic moieties that interacted with what has been termed the tolyl pocket [33]. Analysis of the allosteric site led to a prediction that the protein could also support interactions with fragments of mixed hydrophilic–hydrophobic character in that region; simulated dioxothiomorpholine, ketopiperazine, and diazepanone fragments all bound with high affinity. A molecule was synthesized with a decorated diazepanone group that interacted in the tolyl pocket. This molecule had a p38 $\alpha$  IC<sub>50</sub> of 22 nM and an oral PK (rat) of 33 % F (at 4 mg/kg) [33].

---

## 4 Binding Site Analysis Workflow

The fragment-based analysis of a protein to locate and characterize binding sites requires simulations of small molecule fragments with that protein. First probe fragments are used to find consensus sites, where multiple fragment types of diverse character have strong interactions with the protein. Clusters which are not excluded by high affinity waters or highly solvent exposed identify promising sites for ligand binding and are assessed for druggability.

### 4.1 Steps for Binding Site Identification

#### 4.1.1 Select and Prepare Protein Structure

The starting point for computationally analyzing binding sites is a 3D representation of the protein, typically an experimental structure from X-ray crystallography or NMR. In cases where there is no experimental structure, a computational homology model is generated using related proteins as templates [21]. When seeking new sub-pockets or sites (including allosteric sites), or if the experimental structure is *apo* with a potentially collapsed binding pocket, a flexibility simulation using MD or NMA may be performed [21, 24]. The protein structure that is selected for simulation should be, to the best of one's knowledge, a bio-relevant form of the protein.

If an X-ray crystal structure has regions of unresolved density, missing residues are modeled with care [21]. Missing amino acids may lead to errors in the site identification stage, but an incorrect placement of missing atoms could exclude fragments, also causing errors. The appropriate charge and rotor states are assigned to amino acid side chains [21] and the termini are capped to prevent overestimated fragment binding energies. Finally, a constrained minimization is performed to relax the protein structure [21].

#### 4.1.2 Generate and Cluster Data for Probe Fragments

To generate the necessary fragment data, the probe fragments are simulated individually with the protein. These calculations can be run in parallel on a computer cluster for efficiency [13, 18]. As previously described, GCMC simulations of the protein–fragment system provide information on fragment binding poses and their

associated binding energies [17–19]. A fragment simulation of water is also performed to provide a solvation map of the protein [2, 20, 21]. The poses with the best predicted binding affinity for each probe fragment type are overlaid to obtain consensus clusters of the probe fragments [2, 20, 21]. Fragments of different types within a defined cluster radius of each other are grouped together [2, 20, 21].

#### 4.1.3 Analyze and Select Sites

A sensitivity analysis is performed by adjusting the values of the parameters for the number of fragments in a cluster and the cluster radius [21]. Persistent clusters are kept for further analysis. Clusters are excluded based on the locations of tightly bound water [2, 20, 21] that have a *B*-value below a user defined cutoff. Highly solvent exposed clusters are also removed [21]. The remaining fragment clusters typically indicate a true ligand binding site. This fragment-based technique has performed well at identifying experimental binding sites in a number of protein classes [2, 13, 21].

### 4.2 Steps for Binding Site Characterization

#### 4.2.1 Generate and Cluster Data for Characterization Fragments

In the same manner as data is generated for the probe fragments, fragments used for binding site characterization are simulated with the protein. The consensus clusters for this analysis phase contain fragments from a single characterization category and most of the fragment types in that category must bind with good affinity to define a cluster. Clusters are rejected that overlap with tightly bound water.

#### 4.2.2 Analyze and Prioritize Sites

The following criteria are considered when determining the drug-gability of a site. At least three or four characterization clusters of different categories should fall inside a 15 Å radius of the hotspot cluster defining the site [21]. The fragment clusters should be no farther apart than can be spanned with a linker of three or four atoms. When high affinity waters are found in the binding pocket, they are either re-simulated as part of the protein, or the binding affinity of the designed molecule must be favorable enough to compensate for the water's displacement [21, 32]. The risks of designing ligands to binding sites that do not meet these criteria should be carefully weighed.

#### 4.2.3 Initiate Ligand Design

There are a number of ways in which the binding site characterization stage can be useful for ligand design. If the fragment components of known ligands are simulated, the data can be used to clarify binding modes of those ligands [18, 33, 34]. The fragment analysis can be used to calibrate the simulations by comparing fragment data with structure-activity relationship (SAR) information [18, 34]. If the selectivity of a ligand is a concern for the project, a BSA can be performed on the anti-targets for comparison. Additionally, the fragment clusters interacting with anchoring

hotspots are useful starting points for design, characterization clusters indicate what types of interactions are acceptable in each sub-pocket, and the arrangement of the clusters can provide insight into what chemical scaffolds are appropriate for designing ligands.

## 5 Conclusion

Binding site analysis using virtual fragments is a valuable tool for structure-based drug design. This method identifies bio-relevant binding sites that are druggable, as has been demonstrated by the design of orally bioavailable, nM inhibitors for those sites. The utility of this type of binding site analysis ranges from prioritizing sites based on the predicted difficulty of designing “drug-like” ligands to finding novel binding regions for ligand design.

## Acknowledgments

The author thanks the following colleagues for their contributions to the Locus technology and this approach to BSA: F. Guarnieri, Z. Konteatis, T. Fujimoto, F. Hollinger, M. Clark, J. Wiseman, A. Klon, J. Zou, S. Meshkat, G. Talbot, K. Milligan, and W. Chiang. The author also thanks D. Ludington and M. Ringuette for their editing assistance.

## References

1. Hopkins AL, Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* 1:727–730
2. Vajda S, Guarnieri F (2006) Characterization of protein-ligand interaction sites using experimental and computational methods. *Curr Opin Drug Discov Devel* 9:354–362
3. Henrich S, Salo-Ahen OMH, Huang B et al (2010) Computational approaches to identifying and characterizing protein binding sites for ligand design. *J Mol Recognit* 23:209–219
4. Perot S, Sperandio O, Miteva MA et al (2010) Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov Today* 15:656–667
5. Nisius B, Sha F, Gohlke H (2012) Structure-based computational analysis of protein binding sites for function and druggability prediction. *J Biotechnol* 159:123–134
6. Schmidtko P, Souaille C, Estienne F et al (2010) Large-scale comparison of four binding site detection algorithms. *J Chem Inf Model* 50:2191–2200
7. Labute P, Santavy M (2007) Locating binding sites in protein structures. *J Chem Comput Group*. <http://www.chemcomp.com/journal/sitefind.htm>
8. Le Guilloux V, Schmidtke P, Tuffery P (2009) Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* 10:168
9. Halgren T (2007) New method for fast and accurate binding-site identification and analysis. *Chem Biol Drug Des* 2:146–148
10. Halgren TA (2009) Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* 49:377–389
11. An J, Totrov M, Abagyan R (2005) Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol Cell Proteomics* 6:752–761
12. Guarnieri F (2004) U.S. Patent No. 6,735,530. U.S. Patent and Trademark Office, Washington, DC
13. Moore WR (2005) Maximizing discovery efficiency with a computationally driven fragment

- approach. *Curr Opin Drug Disc Devel* 8: 355–364
- 14. Ringe D (1995) What makes a binding site a binding site? *Curr Opin Struct Biol* 5: 825–829
  - 15. Mattos C, Ringe D (1996) Locating and characterizing binding sites on proteins. *Nat Biotechnol* 14:595–599
  - 16. Allen KN, Bellamacina CR, Ding X et al (1996) An experimental approach to mapping the binding surfaces of crystalline proteins. *J Phys Chem* 100:2605–2611
  - 17. Guarnieri F, Mezei M (1996) Simulated annealing of chemical potential: a general procedure for locating bound waters. Application to the study of the differential hydration propensities of the major and minor grooves of DNA. *J Am Chem Soc* 118:8493–8494
  - 18. Clark M, Guarnieri F, Shkurko I et al (2006) Grand canonical Monte Carlo simulation of ligand-protein binding. *J Chem Inf Model* 46:231–242
  - 19. Clark M, Meshkat S, Wiseman J (2009) Grand canonical free-energy calculation of protein-ligand binding. *J Chem Inf Model* 49: 934–943
  - 20. Konteatis ZD (2010) In silico fragment-based drug design. *Expert Opin Drug Discov* 5:1047–1065
  - 21. Konteatis ZD, Klon AE, Zou J et al (2011) Computational approach to de novo discovery of fragment binding for novel protein states. *Methods Enzymol* 493:357–380
  - 22. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol* 280:1–9
  - 23. Grimme D, Gonzalez-Ruiz D, Gohlke H (2012) Computational strategies and challenges for targeting protein–protein interactions with small molecules. *Physico-chemical and Computational Approaches to Drug Discovery*. Royal Society of Chemistry, London, UK
  - 24. Klon AE, Konteatis Z, Meshkat SN et al (2011) Fragment and protein simulation methods in fragment based drug design. *Drug Dev Res* 72:130–137
  - 25. Gupta A, Gupta AK, Seshari K (2009) Structural models in the assessment of protein druggability based on HTS data. *J Comput Aided Mol Des* 23:583–592
  - 26. Hajduk PJ, Huth JR, Fesik SW (2005) Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* 48:2518–2525
  - 27. Huang N, Jacobsen MP (2010) Binding-site assessment by virtual fragment screening. *PLoS One* 5:e10109
  - 28. Nayal M, Honig B (2006) On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins* 63:892–906
  - 29. Weisel M, Proschak E, Kriegl JM et al (2009) Form follows function: shape analysis of protein cavities for receptor-based drug design. *Proteomics* 9:451–459
  - 30. Krasowski A, Muthas D, Sarkar A et al (2011) DrugPred: a structure-based approach to predict protein druggability developed using an extensive nonredundant data set. *J Chem Inf Model* 51:2829–2842
  - 31. Schmidtke P, Barril X (2010) Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J Med Chem* 53:5858–5867
  - 32. Clark M, Meshkat S, Talbot G et al (2009) Developing technologies in biodefense research: computational drug design. *Drug Dev Res* 70:279–287
  - 33. Moffett K, Konteatis Z, Nguyen D et al (2011) Discovery of a novel class of non-ATP site DFG-out state p38 inhibitors utilizing computationally assisted virtual fragment-based drug design (vFBDD). *Bioorg Med Chem Lett* 21:7155–7165
  - 34. Ludington JL, Fujimoto TT, Hollinger FP (2004) Determining partial atomic charges for fragments used in de novo drug design. 228th ACS national meeting, Philadelphia, PA (Poster)

# **Part IV**

## **Case Studies**

# **Chapter 13**

## **Fragment-Based Design of Kinase Inhibitors: A Practical Guide**

**Jon A. Erickson**

### **Abstract**

Fragment-based drug design has become an important strategy for drug design and development over the last decade. It has been used with particular success in the development of kinase inhibitors, which are one of the most widely explored classes of drug targets today. The application of fragment-based methods to discovering and optimizing kinase inhibitors can be a complicated and daunting task; however, a general process has emerged that has been highly fruitful. Here a practical outline of the fragment process used in kinase inhibitor design and development is laid out with specific examples. A guide to the overall process from initial discovery through fragment screening, including the difficulties in detection, to the computational methods available for use in optimization of the discovered fragments is reported.

**Key words** Fragment-based drug design, Kinase inhibitors, X-Ray crystallography, Fragment screening, Computational chemistry, Structure-based drug design, De novo design

---

### **1 Introduction**

Protein kinases are a class of enzymes involved in important cellular roles, specifically, signal transduction pathways in the regulation of proliferation, differentiation and survival. Over the last 20 years, kinases have become a major target for drug design in many indications, but particularly in oncology and inflammation. Fragment-based drug design (FBDD) emerged in the mid-1990s with Fesik's groundbreaking SAR-by-NMR work and has since blossomed into a major strategy for drug discovery and development. This review provides a general overview of the concepts, techniques, and application of FBDD to kinase inhibitor discovery. Both kinase inhibitor design and fragment-based design have been reviewed extensively, so this is not intended to be a fully comprehensive review of each area, but rather a look at the practical aspects of FBDD of kinase inhibitors. Due to the multidisciplinary nature of FBDD, a survey of biochemical, biophysical, and virtual fragment screening, assembly of cassettes, strategies and techniques in design, and a few examples are highlighted.

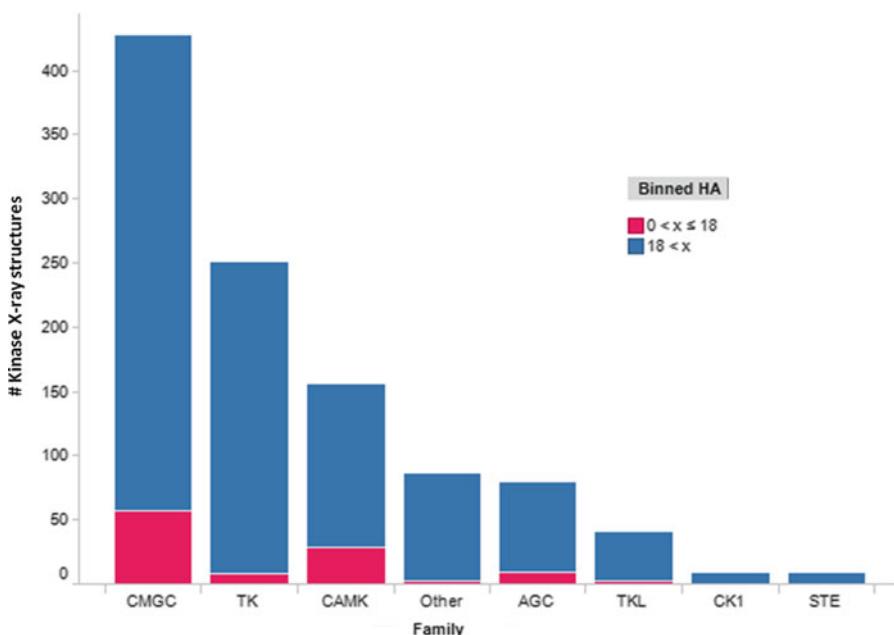
### 1.1 Kinase Overview

Kinases are one of the largest families of proteins in the human genome, consisting of 518 protein kinases and at least 20 lipid kinases [1]. In fact, they are encoded by approximately 2 % of eukaryotic genes. Protein kinases catalyze protein phosphorylation, i.e., the transfer of the  $\gamma$ -phosphoryl group of adenosine triphosphate (ATP) to the hydroxyl group of a tyrosine, serine, and threonine residues in protein substrates. This transfer is part of a vast signal transduction pathway that dictates the regulation of cells. Conversely, their deregulation leads to many disease pathologies. Specifically, human malignancies are associated with activated protein or lipid kinases or inactivated phosphatases. This occurs for many reasons, for example, mutations, chromosomal rearrangements, and/or gene amplification. As such kinases now represent almost one third of drug targets currently being examined by pharmaceutical companies and academic researchers [2]. While their primary therapeutic indications are cancer and inflammatory diseases, kinases are also under investigation as targets for diabetes, infectious diseases, cardiovascular disorders and cell growth and survival. Over the last 10–15 years, several kinase inhibitors have been approved for therapeutic use. Specifically, 11 have been approved for cancer of the 20 overall kinase inhibitor drugs that are on the market. Furthermore, over 150 additional kinase inhibitors are currently in various stages of clinical trials. It is estimated that about 50–70 % of current cancer drug discovery efforts in industry and academia are focused on protein kinase inhibitors. Of the over 500 human kinases, a relatively small number of them are the primary target for the current kinase drugs. These include the tyrosine receptor kinases, EGFR, ERBB2, VEGFR, Kit, PDGFR, the nonreceptor tyrosine kinases ABL and SRC, and one Ser/Thr-specific kinase, the atypical protein kinase, mTOR. In all but the last case, these drugs bind in the ATP pocket. Kinase inhibitor drugs are just now emerging, but their discovery and development still faces many issues such as resistance through mutation, selectivity, a limited number of inhibitor chemotypes, in addition to the need for making inhibitors potent enough to compete with the cellular millimolar ATP concentrations [3]. As such, almost all drug discovery and optimization strategies have been applied to kinase inhibitors, including structure- and fragment-based methods.

### 1.2 Kinase X-Ray Structures

Protein kinases are often classified on the highly conserved sequence similarity of their catalytic domains. Manning et al. subclassified them into eight major groups, namely, AGC, CAMK, CK1, CMGC, STE, TK, TKL, and other using hierarchical clustering techniques and this grouping has been largely taken up as the standard classification of the protein family [4]. Additionally, kinases and particularly their catalytic domain have been very amenable to X-ray crystallography, allowing a detailed analysis of their three dimensional structures. The structures of protein kinases have been

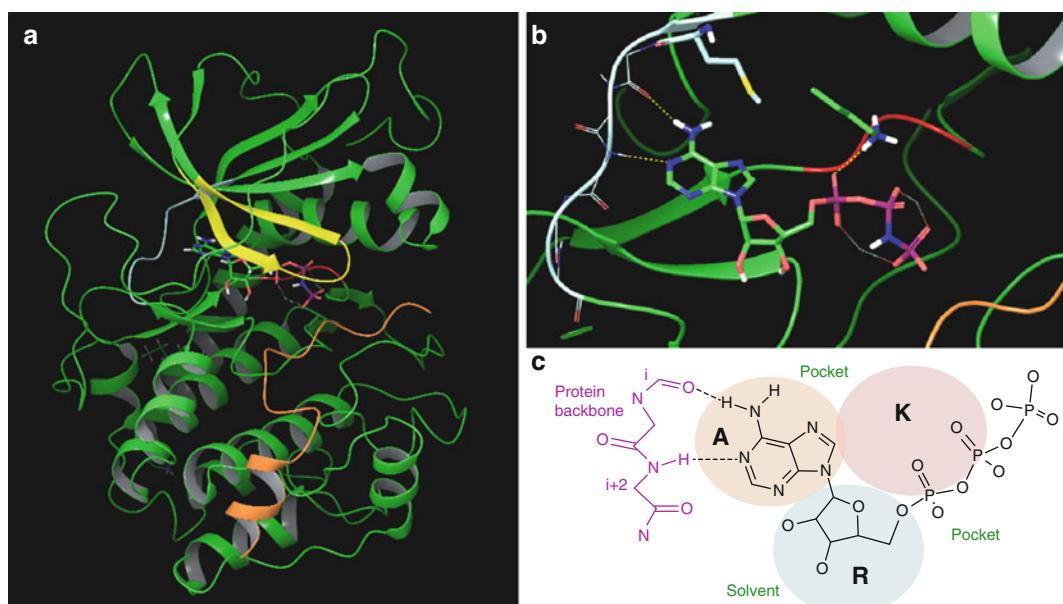
extensively reviewed [5–7] and analyzed, especially with respect to binding of ligands to the ATP pocket [8–10]. Among the 518 kinases, it has been estimated that about 190 unique human kinase structures have been solved encompassing over 1,200 individual kinase ligand co-complex structures [10]. These numbers agree with a similar annotated and aligned database maintained internally at Eli Lilly [11]. For example, we have annotated ~2,200 human protein and lipid kinase X-ray structures from the PDB with some sort of bound ligand, and ~1,600 of those ligands have a unique chemical structure (not including internal X-ray structures). This type of annotation is very useful in the structure-based design and optimization of kinase inhibitors. In fact, ~200 of the unique ligands bound to kinases in the PDB could be considered fragments ( $<= 18$  heavy atoms) in our annotated database. The numbers are a little higher than those found in an internal analysis of the KLIFS database [10] that showed ~1,000 unique ligand structures and approximately 100 fragments. The KLIFS database, however, does not include non-protein kinases or kinases with non-kinase domains. For a breakdown of the distribution of the kinase structures with unique bound ligands refer to Fig. 1. This histogram shows the large number of X-ray structures available from each subfamily. In the 10 years since first publishing the kinase annotated database, the number of kinase X-ray structures has gone up dramatically. At that time, only 38 kinase structures were available. In the



**Fig. 1** Histogram showing the number of X-ray structures of unique ligands bound to kinases and grouped according to kinase subfamily. Fragments, ligands with less than or equal to 18 non-hydrogen atoms, are colored in *pink*, while the larger ligand numbers are represented in *blue*

intervening decade, the number of kinases represented in the PDB has increased five times. At the same rate, all 518 identified kinases will have representative structures in the PDB in the near future. Additionally, the proportion of the fragment structures to larger ligands is highlighted. These kinase structures represent a wealth of data for subsequent fragment-based design. Examples of this use are given in the case study section of this guide (*vide infra*).

The kinase catalytic domain has a bilobal structure with the N-terminal lobe, consisting of mainly  $\beta$ -sheet, connected to the  $\alpha$ -helical C-terminal lobe through a “hinge” strand. ATP binds to this hinge region in a cleft between the two lobes. Figure 2 shows the general overall structure of a kinase, in this case, AMP-PNP (i.e., a non-hydrolyzable form of ATP) bound to cyclic AMP dependent protein kinase A, PKA [12]. The protein is displayed in ribbon format with the N-terminal domain at the top of Fig. 2a. The mostly  $\beta$ -sheet lobe is connected through a “hinge,” (shown in cyan) to the lower C-terminal, which can be seen to consist of mainly  $\alpha$ -helices. AMP-PNP is shown located between the two lobes, located where the cofactor, ATP, binds. Lying over the adenine and ribose rings of ATP is a conserved hairpin loop (in yellow) called the glycine-rich, G- or P-loop. This loop is very flexible and typically changes conformation in response to the bound ligand. In addition, this structure, 1cdk, has a peptide bound in the protein



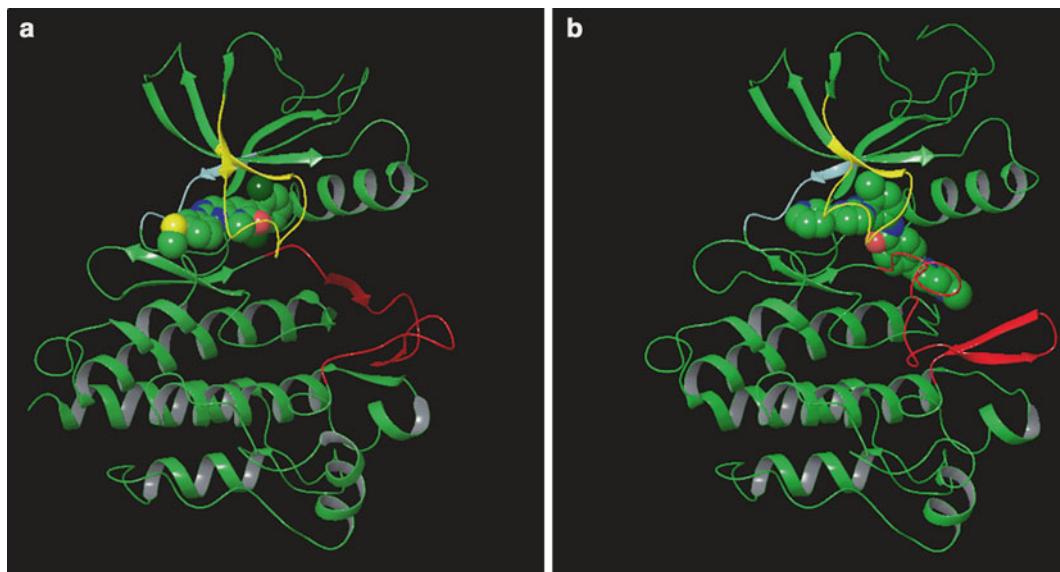
**Fig. 2** (a) X-Ray structure of AMP-PNP bound to PKA shown in ribbon format (PDB code 1cdk). The hinge region is colored in *cyan*, the glycine-rich loop in *yellow*, and the substrate in *orange*. (b) ATP binding pocket of AMP-PNP bound to PKA displaying hydrogen bonds to the hinge region in *yellow*. (c) Annotated schematic of the ATP binding pocket labeled for the various subpockets. A for the adenine, R for the ribose, and K for the lysine subpockets

substrate site (colored in orange) showing the close proximity of where the  $\gamma$ -phosphate of ATP is in relation to the substrate residue that is phosphorylated. In Fig. 2b, a focused view of the ATP pocket is shown. The characteristic hydrogen bonds made by the adenine ring of AMP-PNP to the hinge backbone are highlighted. Also the gate-keeper residue, which is a methionine for PKA, is shown. This residue varies in many kinases and thus provides a “selectivity pocket” which has been exploited in many inhibitor designs [13]. The catalytic lysine in the back-pocket, a conserved residue in all kinases, is also shown. These various regions in the ATP pocket have been extensively explored and described. Figure 2c shows a schematic of the pocket using the annotation according to Liao [9]. Information on the groups that can be accommodated and the interactions available in the various areas in the ATP binding site are critical for structure- and fragment-based design.

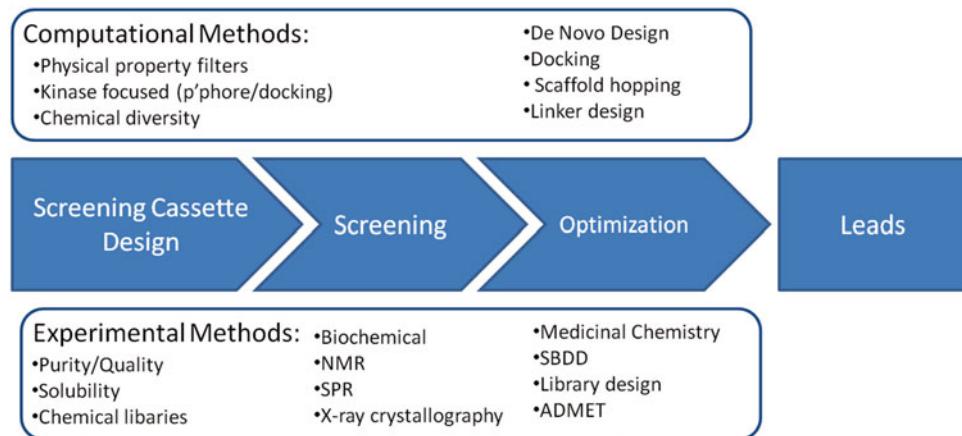
Besides the G-loop, other portions of the ATP binding pocket are conformationally flexible, and thus very sensitive to inhibitor binding. Both the DFG- and G-loop can change in response to the inhibitor, creating a challenge and opportunity in ligand design. In the conserved portion of the back pocket there is a small loop made up of an aspartate–phenylalanine–glycine or DFG motif (red in Fig. 2). This loop typically exists in an “in” structure (as shown in Fig. 2), but depending on the activation state and/or the type of inhibitor bound, it can “flip” into an “out” conformation. In the “out” conformation another pocket is formed, from the movement of the Phe residue, which has been exploited for inhibitor design. There is, however, a downstream effect on the activation loop of the kinase from this movement. The DFG-out conformation can perturb the activation loop into an inactive conformation, similar to that of an unphosphorylated kinase. An example is shown in Fig. 3. The X-ray structure of Abl kinase bound to an inhibitor that does not change the activation loop is shown in Fig. 3a with the activation loop highlighted in red [14]. On the other hand, when imatinib (drug known as Gleevec) binds, it induces or traps an inactive conformation with the DFG loop in an “out” conformation which forces the activation loop into a conformation that blocks the substrate binding pocket (Fig. 3b). Due to these large conformational changes, selecting the appropriate X-ray structure to use for kinase inhibitor design can be critical.

### 1.3 Fragment Concepts

Since Fesik’s seminal SAR by NMR work in 1996, the screening and optimizing of weak affinity compounds with low molecular weight for drug lead generation has mushroomed [15]. In the intervening time, the strategy now termed as fragment-based ligand discovery (FBLD) or fragment-based drug design (FBDD), has become a common tool in the pharmaceutical industry [16]. As the FBLD field has developed, it has converged on some common concepts



**Fig. 3** X-ray structures of Abl kinase in the active and inactive form in ribbon format with the activation loop highlighted in red. (a) Active form of Abl kinase bound to PD173955, PDB code 1m52. (b) Inactive form of Abl kinase bound to imatinib, PDB code 1iep



**Fig. 4** Overview of the fragment-based ligand discovery process. See text for discussion

and processes over time. The general process for application of FBLD is outlined in Fig. 4 but the screening strategy, make-up of the compound screening library and methods of optimizing the fragments differ according to the target and goals of the effort [17]. The overall process doesn't differ significantly from that used in the discovery of more typical size lead compounds. The assembly of screening collections and the methods of detection, however, can be quite different. These methods will be briefly discussed below.

In addition, analysis of screening results must be reevaluated for fragment screens. Since fragments are likely to have very weak binding affinity, the compounds selected for follow-up cannot be chosen from an absolute potency perspective. This philosophy is the key paradigm shift and value of FBLD. Instead of selecting potent, but often large molecules with poor solubility and other properties, smaller fragments with improved physical properties but weaker potency are considered. Fragments can be defined in various ways, but in general, a molecular weight or heavy atom count metric is used. A variant of Lipinski's "rule of 5" [18] was proposed where fragments are defined to have a molecular weight less than 300 Da,  $C_{log P} < 3$ , the number of hydrogen-bond donors/acceptors  $< 3$  each, and  $\leq 3$  rotatable bonds [19]. In order to evaluate these often low affinity compounds the concept of atom efficiency, that is, the amount of free energy of binding per atom in a molecule, was introduced and turns out to be very useful for FBLD. Ligand efficiency (LE), first published by Hopkins, is defined as the free energy of binding divided by the number of heavy (i.e., non-hydrogen) atoms in a ligand [20]. This effectively normalizes the potency of a compound with respect to its size. Selecting compounds with better ligand efficiencies as starting points for lead optimization may be preferable to a potent, but large compound that has a greater potential for poor ADME properties [21]. Although the estimated maximal binding energy per atom is 1.5 kcal/mol, compounds with LE's of ~0.3 is a good benchmark to attain in screening and to maintain during optimization. Atom efficiency analysis can be extended to functional groups as well. Verdonk and Rees showed the use of group efficiency in a fragment optimization of inhibitors of the kinase, PKB [22, 23]. Here the efficiency of a group of atoms is compared to the ligand without a group at the position in question or to a common analog in a series. This allows comparison of functional groups in terms of atom efficiency during a structure–activity relationship (SAR) study. Another very useful concept that is aimed at normalizing compounds for properties is ligand lipophilicity efficiency (LLE) [24]. This metric consists of the difference between the  $pIC_{50}$ ( $-\log IC_{50}$ ) and its  $\log P$ . Since lipophilic compounds can bind through a mechanism driven by the entropy gain that occurs by simply displacing weakly bound waters, LLE can help to focus on compounds with a balance of entropic and enthalpic driven binding energies. Values between 3 and 5 are considered a good range to attain for optimization in order to avoid overly hydrophobic molecules. These terms and others like them can be important for kinase discovery, as ATP competitive inhibitors often tend to have low solubilities due to their flat nature and hydrogen bonding functionality. FBLD emphasizes the discipline to adhere to atom efficiency concepts throughout optimization in order to maintain the good properties inherent in the initial fragment hit.

---

## 2 Methods

### 2.1 Fragment Detection

Fragment identification for a particular target can be challenging using traditional biochemical screening methods. Due to the weak affinity of most fragments, biochemical screening paradigms at concentrations up to millimolar concentrations can lead to many false positives. For this reason typical fragment identification protocols employ an additional biophysical detection method to identify hits and/or to confirm true positives from high concentration biochemical screens. Nuclear magnetic resonance (NMR), surface plasmon resonance (SPR), isothermal titration calorimetry (ITC) and X-ray crystallography are examples of successful methods used for kinase FBLD [25, 26]. Since these methods have been extensively reviewed, only a brief summary of each method is given in this chapter.

NMR-based methods applied to fragments can be carried out from the ligand or protein perspective [17, 27]. Ligand-based methods have the advantage of working without need for labeled protein and rely on the high sensitivity of NMR for the free and bound states [28]. Ligand-based NMR will also work at much lower protein concentrations than is needed to detect changes in protein resonances. One can observe many changes (NOE, relaxation or magnetization transfer) of the ligand NMR spectrum to indicate whether it is binding or not making it useful for screening. The saturation transfer difference (STD) experiment is commonly used due to the ability to use very low ratio of protein to ligand concentrations. This method can also be used in a competition-like experiment to displace a “spy” molecule to confirm binding sites. Ligand NMR, however, does not give any binding information so is typically followed up with X-ray crystallography. To get binding location information, one can also detect changes in protein <sup>15</sup>N chemical shifts upon ligand binding in HSQC or TROSY spectra. This data can provide not only binding confirmation but also information on where the fragments bind. The requirement of labeled proteins and the limitation on the size of the protein does constrain the applicability of the protein experiment.

Another versatile biophysical screening method involves surface plasmon resonance (SPR) [29]. The method typically involves immobilization of the protein on a metal (typically gold or silver) surface that the ligands/fragments are passed over. Oscillations in the electrons (plasmons) excited by polarized light can be detected and related to the mass change. Bound and unbound molecules can then be quantified in a binding and subsequent rinsing experiment yielding kinetic and thermodynamic data, i.e., on/off rates and thus dissociation constants. Recent advances in SPR detection, allow the method to be used in a high throughput format, permitting the screening of large fragment libraries.

Isothermal titration calorimetry (ITC) is another technique that has been used to identify fragments [30]. This technique directly measures the heat of binding, allowing the calculation of thermodynamic terms,  $\Delta H$  and  $\Delta S$ . These terms provide valuable information in the selection of compounds to pursue and also in their optimization. ITC is primarily used as a secondary screen, as the amount of protein required and lack of high throughput formats preclude it as a method to screen large libraries of compounds.

Perhaps the most information rich fragment identification method is X-ray crystallography. It not only allows detection of fragments, but directly gives the binding mode (*see Note 1*). Some of the earliest fragment work was carried out with X-ray crystallography. The seminal work by Ringe and coworkers used organic solvents in crystallography to map binding pockets of a protein [31]. In most cases, however, X-ray crystallography is not amenable to screening large libraries and thus is used mostly to follow-up confirmed hits found from other detection methods [32]. That said, several groups have developed methods to use X-ray crystallography in a direct screening fashion [33, 34]. These are mainly accomplished through soaking of cocktails of fragments/ligands into preformed crystals. Astex, SGX (currently Eli Lilly), and other companies have used this methodology to identify and optimize fragments. Not all proteins have been amenable to this technique, but it has been shown to particularly useful for kinases.

## 2.2 Computational Techniques

Like biophysical methods, computational and informatics technology plays an integral part of fragment-based drug discovery. Computational tools are used throughout the process, from initial screening to compound design and finally lead optimization (*see Note 2*). Here a brief kinase-focused overview of computational fragment methodologies is outlined.

Many of the detection methods for fragments relies on low throughput methods so libraries of compounds to screen must be assembled in order to give the best chances for a positive outcome, i.e., a variety of starting points with good ligand efficiency (*see Note 3*), physical properties and an opportunity for optimization. Furthermore, the need for highly soluble compounds is essential due to the high concentrations needed to detect binding. Fragment screening libraries can be constructed as general screening pools to represent chemical diversity or as focused library on a particular target. Many examples of general fragment screening libraries have been published and reviewed [35]. An example is the one described by Astex [36] but other similar libraries have been used by other companies [35]. Astex began by identification of rings and functional groups found in known drugs to create an initial virtual database that was expanded to topological equivalents and enumerated with other atoms and single substitutions of favored functional groups [36]. A database of ~4,000 compounds

was produced of which ~400 were available for purchase. This work was somewhat based on the scaffold generalization paradigm used at Vertex [37]. They used a similar approach for an NMR screening library that was based on a small set of fragments designed to represent a diversity of shapes [38]. In addition to general libraries, target focused libraries have also been created and used for kinase fragment screening. For example, Astex used docking and a hinge-specific pharmacophore to select fragments for a kinase-biased library (*see Note 4*) [36]. In a similar fashion, a kinase pharmacophore based on hinge binding in kinase X-ray structures was used to create a kinase-biased NMR screening library [39]. The “privileged-structure” concept was used to greatly enrich screening results for kinases by workers at Vertex [40]. Using a “kinase-likeness” parameter developed from scaffold or “framework” analysis of kinase inhibitors to select compounds for screening, gave up to a fivefold enrichment in kinase screening experiments. Docking-based methods have also been used to focus screening sets on kinase targets [41]. More recently QSAR and machine learning methods have yielded highly enriched screening cassettes for kinases. For example, support vector machine [42] and Bayesian [43] models have shown high enrichments in kinase screening. Sutherland et al. showed that naïve Bayes models built from kinase and other target family inhibitor fragment fingerprints could predict potency quite well and thus would be another method to select compounds for kinase fragment screening [44].

Computational methods are also quite useful in specific design and optimization of fragments, using many structure-based design methods. Selection of kinase fragments for optimization typically begins with an X-ray structure of the small compounds bound to the kinase of interest, as docking of fragments to estimate the binding mode of small fragments has been a challenge. Even though fragment docking should be somewhat simplified due to a lack of conformation degrees of freedom, inaccuracies and molecular weight dependence of most scoring functions are often unable to distinguish the correct pose as defined by the X-ray structure. This concept was examined in a recent study that systematically compared fragment with drug-size molecule docking [45]. They concluded that docking accuracy of fragments did not differ significantly from that of docking larger molecules. Interestingly, the root cause for accuracy did differ. It was observed that scoring functions were the main problem with fragments, while sampling of the ligand conformation caused inaccuracies with larger molecules. The overall cross-docking result for fragments, however, showed low accuracy, ~40 % success rate, using a 1.5 Å RMSD successful docking criterion. In another recent study, cross-docking results with Glide show somewhat better results, ~60 % accuracy, but their definition of a fragment was larger and they used a 2.0 Å RMSD to define a successful pose. An interesting prospective analysis of

PIM1 kinase docking highlighted some of the value and pitfalls of fragment docking [46]. Analysis of hits from a fragment screening experiment that used docking to select the compounds going into the screen, showed some interesting results. Docking was able to identify several efficient fragments, but in two of the five cases discussed, the docking pose didn't agree with the subsequent X-ray structure. This was mainly due to water positions and the misplacement of hydrophobic portions of the fragments. These observations highlighted the critical need for an X-ray structure in FBDD if available. Docking, however, was able to reproduce the important hydrogen bonding patterns that were useful in the design of new compounds. The take home learning is that docking can be useful, but care should be taken in the use of the results and iterative X-ray crystallography should be used whenever possible. Other computational methods have been employed for fragment placement besides docking. Karplus's MCSS [47] protocol was the pioneering fragment placement and binding site mapping algorithm actually predating the advent of experimental fragment-based methods. The multiple copy simultaneous search method (MCSS) floods a protein binding site with thousands of copies of a fragment followed by simultaneous minimization of the copies. The most energetically favorable positions of the fragments can be used to provide guidance on the position and types of interactions to be used for structure-based design. Another fragment placement method also involves simulation of many fragments in the context of a protein. The Grand Canonical Monte Carlo (GCMC) method solvates a protein in a fragment bath and calculates the ligand–protein binding modes and their binding free energies by annealing [48, 49]. Fragment stability at different energy levels can point to those that are most viable for optimization. Again the position and interactions that these fragments make with the protein can help guide structure- and fragment-based design by direct linking or inference.

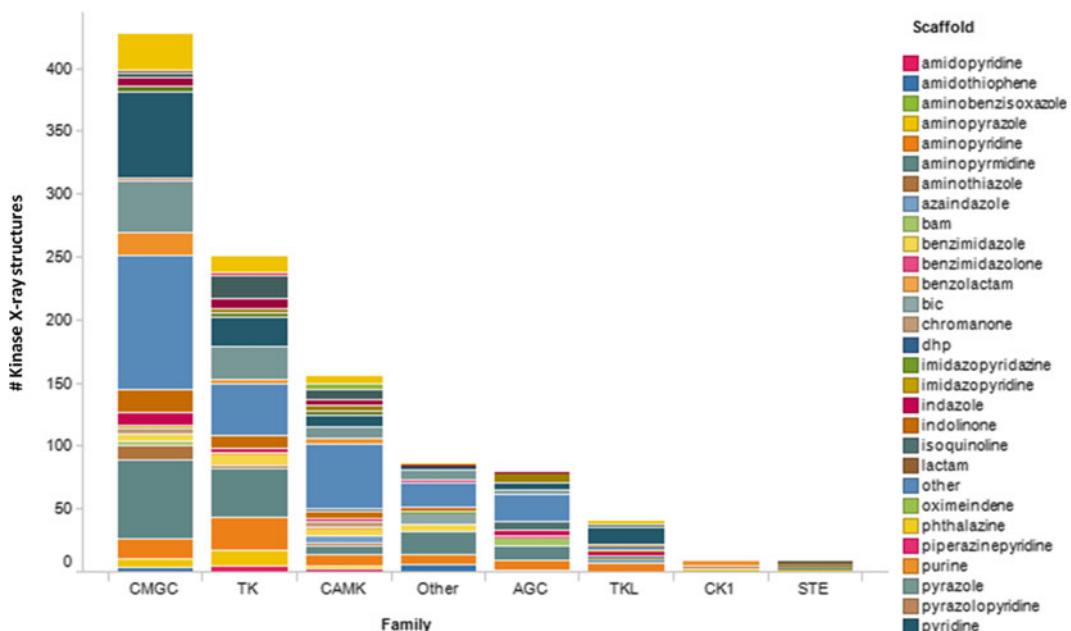
Fragment-based ligand design, after selection of the original fragment placement in the binding pocket, typically employs a linking or a growing strategy. In the latter case, one typically begins with a bound fragment from an X-ray structure or a computationally placed fragment. Positions on the fragment that can be substituted and yield vectors able to access pockets or additional interactions to the protein are then identified. In a linking strategy, substitutable valences on two fragments in the binding pocket are identified followed by finding an appropriate linker group to combine the two fragments. This was strategy was initially carried out in the SAR-by-NMR experiments that launched fragment-based methods [15]. For both strategies, computational techniques have been developed and have evolved over the years, mostly under the term “de novo design” methods. In fact, most of the techniques used to place, link, grow or otherwise optimize fragments come

from structure-based de novo efforts. One of the earliest fragment placement and linking methods was pioneered by Bohm [50, 51]. The LUDI algorithm automatically placed fragments in a protein binding site using molecular interaction rules derived from the PDB and CSD databases. In addition, small fragment databases were provided to link fragments together and also for growth from a fragment to pick up additional molecular interactions. Other linking methods were also developed. The HOOK program was developed to link fragments generated using the MCSS method [52]. In another approach, a functional group linker was developed by Bartlett's group [53] to find optimal rings to position two fragments based on their respective linking vectors. The initial implementation of CAVEAT has been updated and augmented with the use of substituted rings [54]. This idea was further expanded by implementing a more general vector alignment linking protocol using experimental or computed ring conformations to replace scaffolds in a bound ligand [55]. The method, called RECORE, allows linking and growing plus an option to use the protein structure as a constraint along with a scoring function to rank order ideas. Besides linking, LUDI was also able to grow from a fragment in an attempt to satisfy extra potential interactions in a protein binding pocket. Another seminal development in structure-based fragment growth was the work of Bohacek and McMMartin [56]. The GrowMol algorithm begins with a pre-generated grid enclosing the binding pocket with a heuristic complementarity potential encoded on each grid point that attempts to capture the chemical environment of the local area of the pocket. The available growth points are identified on the fragment or molecule and atoms and functional groups are added in a stochastic manner (metropolis sampling) and scored for complementarity for retention. The method has the added ability to grow rings and macrocycles. In another early approach, Dean and coworkers initially utilized placement of fragments and concomitant growth through the addition of predefined groups in a protein binding pocket [57, 58]. They used a simulated annealing algorithm to drive the optimization allowing movement of the new structures in the pocket, thus effectively sampling chemical, conformational, and positional spaces simultaneously for new molecules. The program, Skelgen, has been improved to add key innovative steps to increase its usefulness. The addition of a fragment library generated from the RECAP rules [59], which disconnects molecules according to retrosynthetic rules, helped Skelgen produce more synthetically feasible molecules. The ability to incorporate side chain flexibility of the protein during design to accommodate induced fit was also added [60, 61]. CDK2, was used as an initial test case for Skelgen, showing its usefulness for kinases [62]. A final innovative strategy for structure-based fragment or de novo design involved a chimera-like approach utilizing the X-ray structures of

ligands bound to the same protein or family of proteins [63]. BREED matches bonds using a geometric (distance and angle) criteria in ligands overlaid by aligning their protein X-ray structures. Once the bonds are matched, molecular fragments are recombined to produce hybrid molecules. This method was also demonstrated using two kinases, p38 Map kinase and CDK2, and showed an ability to generate kinase inhibitors that fell under the scope of a p38 inhibitor patent. An exhaustive recount of all the specific de novo strategies for fragment-based design has not been given here, but rather an attempt to highlight the main strategies. Many groups have taken these basic ideas, augmented and improved them through expanded fragment databases, scoring function development and increased conformational sampling. In practice, many of these techniques are available through commercial software packages. Examples are the fragment-based capabilities from within the Schrodinger Inc. [64] or Chemical Computing Group's (CCG) [65] suite of programs. These groups have implemented several of the fragment-based methods discussed herein and integrated them into useful graphical interfaces. Schrodinger's docking program, Glide, can be used for fragment placement and docking-based selection of fragment libraries to screen. They also have routines for structure-based combinatorial library expansion of fragments and core-hopping ability (CombiGlide) with supplied fragment databases and all integrated into their Maestro interface. CCG's fragment tools are even more extensive. They have a specific fragment menu in their MOE interface that allows access to almost all of the fragment elaboration and de novo design strategies discussed. In particular, MOE contains linking and growing methods that allow facile fragment elaboration with included fragment libraries. They have provided access to an implementation of the RECAP rules in order to allow users to create their own fragment databases. Fragments can be evolved with either an X-ray structure or pharmacophore model using a number of strategies with their tools. Fragment replacement with bioisosteres, medicinal chemistry rules or through an implementation of the BREED algorithm can also be carried out. Finally, CCG has created an annotated kinase X-ray database, which allows a by-residue analysis and comparison of kinase ATP binding sites as well as annotation of the state of the DFG loop conformation (*see Note 5*). As with the KLIFS database, this annotation is very useful for design and optimization of both potency and selectivity. Thus, in practice, fragment-based design can be carried out readily given the availability of a X-ray structure of a fragment bound to the kinase of interest using already existing structures and computational tools.

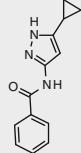
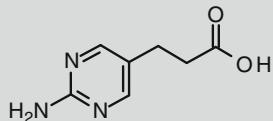
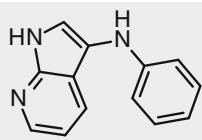
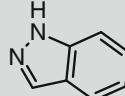
The starting place for most applications of kinase fragment-based design is an X-ray structure of a small ligand bound to the kinase of interest. This fragment typically comes initially through screening (*vide supra*) or mining of the numerous already available

kinase-fragment X-ray structures. As previously highlighted, kinases are particularly amenable to X-ray crystallography and as such provide an ideal target family for FBDD and FBLD. Many annotated kinase X-ray structure databases have been created to help with inhibitor design. In Lilly's version [11], we have created tools to identify common scaffolds found as hinge binders in the kinase-ligand co-complex X-ray structures. All of the kinase structures are aligned in a common frame of reference so the ligands bound can be readily compared according to a binding mode. Figure 5 shows a partial classification of the structures from the KLIFS database using our scaffold classification method. Many common scaffolds are represented in ligands bound throughout the various subfamilies of the kinome. Furthermore, we (and others) have recognized that most of the kinase hinge interacting scaffolds, bind in a similar way across the various kinases. There are many exceptions to this observation, but there is enough commonality in scaffold-binding mode that it can be used in a reasonably predictive fashion. In other words, given a particular mode that a kinase inhibitor scaffold binds to the hinge of one kinase, there is a good chance that it will have a similar binding interaction with the other kinases. We have used this to qualitatively predict binding modes for fragment- and structure-based design of kinase inhibitors [66, 67]. In fact, in many cases a hinge-binding fragment X-ray structure can be predictive of a more elaborated



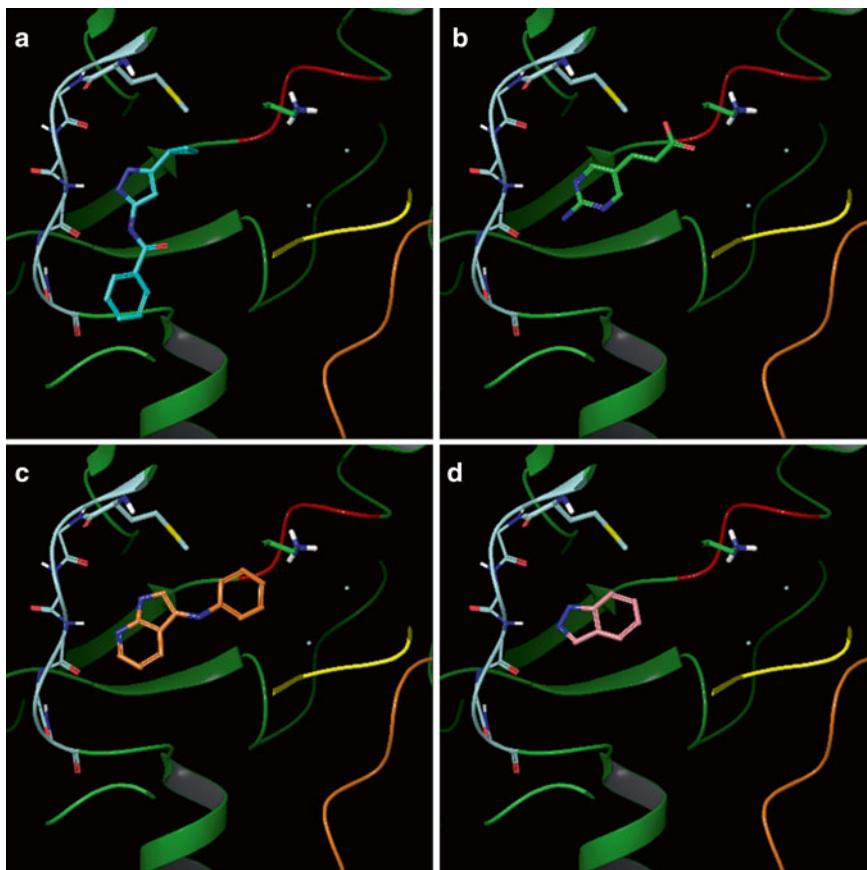
**Fig. 5** Histogram showing the number of X-ray structures of unique ligands bound to kinases and grouped according to kinase subfamily. Here the ligands are *colored* by the scaffold as classified with an in-house method. See text for further description

**Table 1**  
**Analysis of the conservation of kinase ligand binding modes**

Scaffold	Number of X-ray structures (fragment)	Number of X-ray structures with common binding mode (fragment)	Representative fragment	PDB code
Aminopyrazole	26 (2)	21 (2)		1vyz
Aminopyrimidine	130 (7)	91 (4)		2xj0
Azaindole	20 (3)	16 (3)		3c4e
Indazole	22 (5)	21 (5)		2vta

The analysis is based on alignment of kinases proteins X-ray structures from the PDB in a common frame of reference

compound with the same hinge binder. For example, four common hinge binding scaffolds (i.e., from data shown in Fig. 5) with both fragment and larger molecules X-ray structures were examined and results shown in Table 1. For each scaffold, a majority of the binding modes were qualitatively the same as the fragment, even though the structures represented by these scaffolds bound are in many different kinases. The binding modes of the four representative fragments are shown in Fig. 6 to highlight the binding mode in relation to the hinge region. Figure 7 shows an overlay of the fragment-containing larger molecule structures on these fragments. Since the overlays are of the kinase proteins and not the ligands, the scaffolds don't exactly align, but have the same relative binding mode as the fragment structure for many ligands. This gave us confidence to utilize predicted binding modes from previously crystallized kinase inhibitors of the same or similar scaffold in library design. This method of predicted binding mode was utilized in the de novo design of several kinase libraries and fragment expansion using the predicted binding modes led to the synthesis of several kinase active libraries [68].

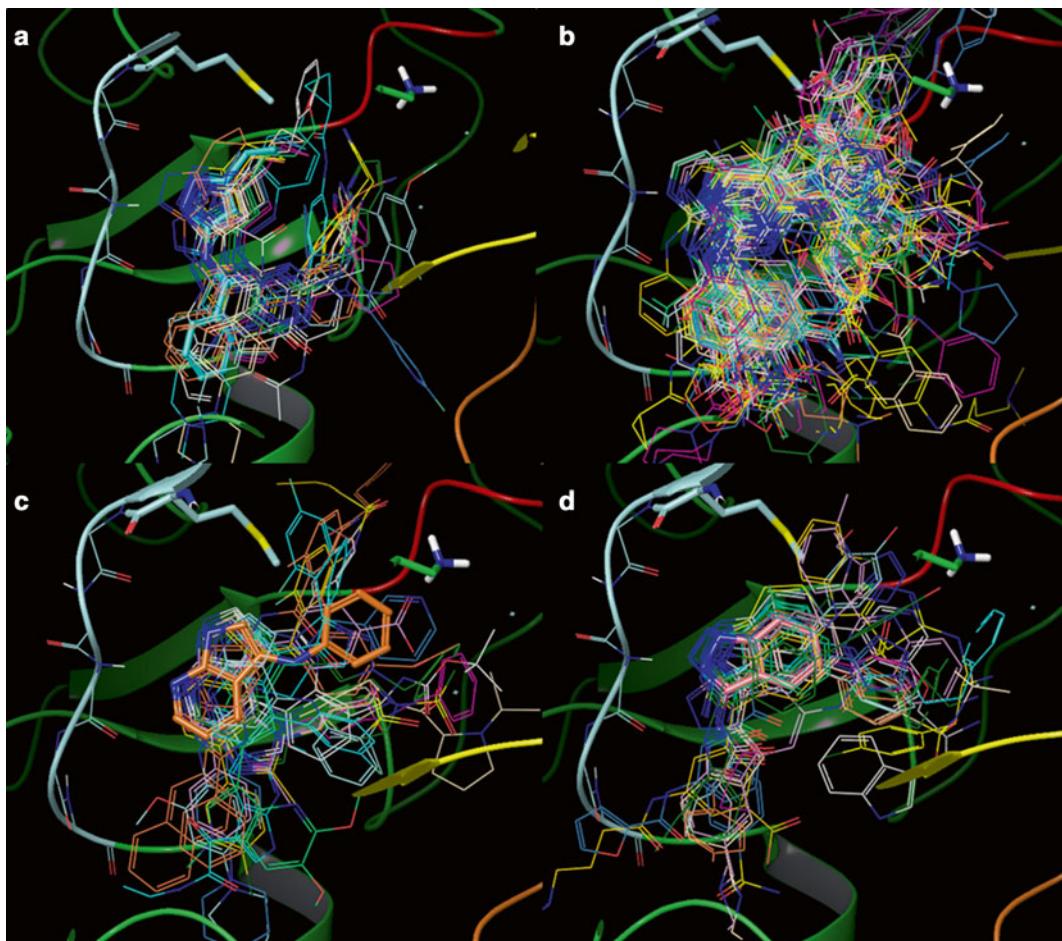


**Fig. 6** X-Ray structures of four fragments from Table 1 bound to various kinases. (a) Aminopyrazole scaffold, PDB code 1vyz. (b) Aminopyrimidine scaffold, PDB code 2xj0. (c) Azaindole scaffold, PDB code 3c4e. (d) Indazole scaffold, PDB code 2vta

### 2.3 Recent Case Studies

Since many excellent reviews on FBLD and FBDD have been published showcasing very successful and elegant practical kinase inhibitor examples [36, 69, 70], here only a few recent reports of the various FBDD strategies are summarized, *see* Table 2. These reports were identified from recent kinase bound fragment structures in the PDB, which utilized FBLD for discovery and/or optimization. They represent a variety of kinases and scaffolds and represent the wide applicability of FBLD and FBDD for kinases.

Astex, one of the pioneering companies for FBLD and FBDD, provides a nice example of the fragment process in a recent report on discovery of CDK2 inhibitors [71]. Using a CDK2 X-ray crystallography soaking system, they screened their fragment library to identify viable starting points for optimization. The library consisted of 500 fragments from general and kinase focused methods. They uncovered over 30 hits, and described evaluation of four. As typical for initial fragment hits each of the four had



**Fig. 7** Overlay of the X-ray positions of larger ligands on the four fragments from Table 1 showing the conservation of binding mode. (a) Aminopyrazoles, (b) Aminopyrimidines, (c) Azaindoles, (d) Indazoles

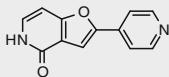
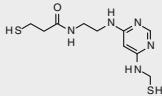
relatively low binding affinities (ranging from ~100  $\mu\text{M}$  to 1 mM IC<sub>50</sub>s) but the hits were very efficient with LE's of 0.37–0.57. The report illustrates the use of X-ray crystallography to grow and optimize three of the four fragments. They nicely emphasize that the key to the efficient optimization is examination of the molecular interactions the fragment (and from all the other fragment structures) displays in the X-ray structures and identification of substitution points that allow growth to form new interactions, often described as “vectors.” For example, they immediately went from mM activity to 7  $\mu\text{M}$  by adding a phenyl group to the amine of the aminopyrazine fragment in order to fill a hydrophobic pocket that was seen in the X-ray structure. In another example, they were able to attain 30 nM potency of a pyrazolopyrimidine fragment by picking up an additional hydrogen bond in the hinge region plus the addition of a piperidine group in the ribose pocket to access

**Table 2**  
**Summary of kinase FBLD examples**

Kinase	Scaffold	Representative fragment structure (s)	PDB code (s)	Strategy	Reference
CDK2	Pyrazine		2vta; 2vth; 2vtm; 2vtj; 2vtr; 2vts; 2vti; 2vtl; 2vtn; 2vto; 2tp; 2vtq; 2vtt; 2vu3	X-ray screening and SBDD optimization	[71]
	Indazole				[71]
	Hydroxynaphthalene				[71]
	Pyrazolopyrimidine				[71]
AuroraA-B	Pyrazole		2w1d; 2w1f; 2w1c; 2w1e; 2w1g	X-ray screening and SBDD optimization	[72]
JAK2	Indazole		3e62; 3e63; 3e64	X-ray screening and SBDD optimization	[73]
PIM1	Benzofuran		3r00; 3r01; 3r04	SPR screening SBDD optimization	[46, 74]
Mps1	Quinazoline		3hmp	High concentration biochemical screening and SBDD optimization with MOE tools	[75]
CHK1	Pyrrolopyrimidine		2wmq; 2wmr; 2wms; 2wmt; 2wmu; 2wmv; 2wmw; 2wmx	High concentration biochemical screening and SBDD optimization with MOE tools	[76–78]
Rho Kinase	Indazole		3vbs	High concentration biochemical screening and SBDD optimization	[79]

(continued)

**Table 2**  
(continued)

Kinase	Scaffold	Representative fragment structure (s)	PDB code (s)	Strategy	Reference
GSK3b	Pyridine		3zrk; 3zrl; 3zrm,	Kinase panel screening and SBDD optimization	[81]
GSK3b & FLT3	Triazole		None	Kinase panel screening and SBDD optimization	[80]
PDK1	Pyrimidine		3qc4	Covalent tethering to screen thiol library followed with SBDD optimization	[82]

another hydrogen bond. The indazole fragment was carried the furthest in the optimization. It was, in fact, morphed through several FBDD iterations into a diamido-pyrazole which was eventually advanced to clinical trials. In another case of FBLD, researchers used a scaffold, specifically, pyrazole-benzimidazole, which was a fragment identified not for the target of interest, Aurora A, but in an X-ray screen for CDK2 inhibitors [72]. Characterization of these compounds against a panel of kinases showed good potency and especially, efficiency, for Aurora A kinase. This is a good example, of transferring scaffolds from one kinase to another. X-ray crystallography of the pyrazole-benzimidazole bound to both kinases showed a similar binding mode to both CDK2 and Aurora A and explained the additional potency of this fragment for Aurora A as the complementarity of this scaffold to Aurora A near the hinge was better than it was for CDK2. Astex was able to take advantage of this complementarity to grow this fragment into a potent and selective compound that was also advanced to clinical studies. During the optimization, they again emphasized the attention to physical properties, specifically size and lipophilicity. The final molecule, for example, featured a cyclopropyl urea group that provided the right conformation and lipophilicity for cell potency and solubility. Another indazole fragment found for JAK2 kinase through X-ray crystallographic screening was recently reported [73]. Using the SGX X-ray crystallographic screening protocol, a fragment library designed to have good physical chemical and drug-like properties, yielded several hits. They selected a bromo-amino-imidazole to follow-up ( $41 \mu\text{M} \text{ IC}_{50}$ ) for a fragment growth

strategy due to its high LE (0.54) and its potential substitution points. Using the X-ray structure, viable vectors for growth were selected, namely the 5- and 6-positions of the indazole, and several analogs were synthesized. A highly atom efficient 5-phenyl substitution was identified and selected over the less efficient 6-position for further elaboration. A very potent 4-t-butyl-sulfonamide substitution on the 5-phenyl ring was designed and yielded very potent binding (78 nM). From the initial fragment and synthesis of a minimal number of analogs (<25), they achieved a 500-fold potency increase. Other screening methods have also been quite successful for FBLD. An example of a fragment campaign using SPR screening was recently reported for PIM1 [46, 74]. In this effort, a 1,800 compound fragment library was screened using SPR at high concentration (75  $\mu$ M) with the hits being followed up with biochemical IC<sub>50</sub> determinations. Several active fragments were found, but a few 2-carboxylic acid benzofuran analogs were particularly interesting due to their high LE values and novelty for PIM1. This kinase is quite unique in its structure as it has a proline residue in the hinge position whose backbone NH normally acts as a key hydrogen bond donor with the purine ring of ATP and to most kinase inhibitors. This is a case where previous co-crystal structures of the scaffold bound to other kinases will most likely not transfer due to the unique hinge in PIM1. Thus, X-ray structures were critical here for use in growth and optimization of the fragment. The X-ray structure and activity relationship of the benzofuran hits suggested 5-position and 7-position elaboration. A flip in the binding mode of about 180° occurred when the 5-bromo analog was substituted with a 7-methoxy group. In effect the flip overlays the two bromine atoms in the same hydrophobic pocket while allowing the methoxy group to project into an open area. This is an example, where the binding mode can change upon growth, emphasizing the importance of iterative X-ray crystallography for fragment work. Further optimization is described exploring the 5- and 7-position vectors leading to a compound with a 425-fold potency increase. In addition, the physical properties of the fragments were also very important during the growth and optimization. Some of the optimized compounds proved to be very soluble and metabolically stable. The next two examples utilized a similar process of high concentration screening to identify fragments from cassettes built using kinase-based pharmacophore methods in MOE. In the first case, a screening cassette of commercial compounds was constructed with the help of a MOE 3D pharmacophore to focus the fragments for screening against Mps1 kinase [75]. In particular, a set of five different kinase X-ray structures was used to build a pharmacophore model which was subsequently used to filter 15,000 commercial compounds. In addition, the compounds were filtered using general drug-likeness models to comply with good size and physical properties.

The resulting hits were clustered by scaffold into a 160 compound set that was used in the biochemical screening at a 50  $\mu\text{M}$  concentration. Six quinazoline hits were found and followed up with X-ray crystallography from which an Mps1 X-ray structure co-crystallized with a 13  $\mu\text{M}$  hit was solved. Full characterization of the molecular interactions provided data important for further growth and optimization. A second example of the using of high concentration biochemical screening involved the discovery of a CHK1 inhibitor [76–78]. Again using a 3D pharmacophore built with the MOE tools was enlisted to filter commercial databases for an initial screening set. In addition to the kinase pharmacophore, which required three hydrogen bonds to the hinge region, a hydrophobic group in the adenine-ribose pocket and an interaction with the catalytic lysine, they also used a CHK1 X-ray structure as an excluded volume constraint. Hits from the screen were clustered into chemotypes and representatives from each scaffold were selected to yield 361 compounds for high concentration biochemical screening. Of the nine hits that were confirmed by X-ray crystallography, a pyrrolopyrimidine fragment was particularly efficient and thus selected for some initial structure-guided growth. This scaffold was later morphed into an amino-isoquinoline which showed good activity in mouse xenograft models. Instead of screening commercial or in-house databases of compounds, several efforts demonstrated the use of X-ray data to design one-step reactions to generate screening cassettes for FBLD. One example utilizing this strategy focused initially on small library of 4-pyridyl analogs [79]. High concentration biochemical screening identified several weakly active (72–270  $\mu\text{M}$ ), but atom efficient analogs (0.29–0.39) for Rho kinase. Modeling suggested substitution of the 4-pyridyl hinge binder with an indazole which further increased the potency. Optimization ultimately produced potent compounds with promising cell potency against Rho kinase. A similar approach utilizing synthetic libraries combined with FBLD principles was used to identify inhibitors of GSK3 $\beta$  and FLT3 [80]. Here the authors took advantage of click chemistry from which they were able to assemble 1,4- and 1,5-di-substituted 1,2,3-triazole analogs. The 1,2,3-triazole group was designed to bind to the hinge region of the kinases while the various di-substituted regiosomers could explore the pockets of the kinases. Screening of the initial triazole library against a panel of kinases uncovered inhibitors of both GSK3 $\beta$  and FLT3. Another example of panel screening of fragment-like compounds reported was the discovery of thienopyridinone inhibitors of GSK3 $\beta$  [81]. Crystal structures obtained on the first few fragments found and the information obtained from the binding mode regarding the molecular interactions were used to grow the molecule. The first pass optimization consisted of a structure-guided medicinal chemistry approach followed by a computational driven approach to optimize the CNS permeability

properties of the compounds. Virtual libraries were enumerated and filtered with calculated ADME and physical properties. This was followed by docking to models constructed with the initial X-ray structures to select the final molecules for synthesis. In the end they were able to improve properties for blood–brain barrier penetration and increase potency sixfold. Most recent examples of FBLD and of those presented here, describe fragment growth to optimize ligands rather than linking of fragments. This is mainly due to the common use of X-ray structural data, where the binding of multiple fragments in the same structure are not as common as single fragment binding. An interesting alternative approach uses a technique of tethering a fragment covalently to the kinase followed by X-ray crystal structure guided elaboration [82]. In this instance, a cysteine was introduced at position 166 in the ATP binding site of PDK1 kinase and then coupled to a diaminopyrimidine thiol to covalently bind to the protein. The X-ray crystal structure of one of the optimally linked analogs showed that the aminopyrimidine attained a binding mode bound to the hinge region of PDK1. The bound fragment also had a free thiol, so was able to form covalent bonds readily with other thiols with complementary substituents. To optimize the initial aminopyrimidine, the PDK1 protein with the covalently attached diaminopyrimidine thiol was in turn subjected it to a library of 3,000 thiol containing groups to identify optimal groups for binding in the back pocket of PDK1. A pyrimidone substituted analog was found, and an alkyl chain version of the thiol compound found to bind at 200 nM IC<sub>50</sub> without the covalent tether. Exploration of the tether lengths plus aminopyrimidine replacements led to the inhibitors that bound to a DFG out conformation of PDK1, the first demonstrated by X-ray crystallography.

With the continued advancement of fragment detection methods, computational tools and the ever-increasing number of fragment bound kinase X-ray structures, the future of FBLD to discover new kinase inhibitor drugs is very bright.

---

### 3 Notes

1. Fragment-based methods have had markedly good success in the development of kinase inhibitors for drugs. The amenability of kinases to X-ray crystallography and especially the ability to solve fragment bound structures has driven this success.
2. The most challenging aspect for kinase FBLD comes in during the optimization of the fragments. Selectivity, potency, solubility and ADME properties can be very difficult issues to surmount, but with the ability to use iterative X-ray crystallography and computational tools, many examples of successful kinase

FBLD have been noted. In fact, one of the first drugs to originate from a fragment campaign is the BRAF kinase inhibitor, Vemurafenib [83, 84].

3. One reason kinases have been very good targets for FBLD is the high ligand efficiency with which fragments bind to this class of proteins. Highly efficient fragments are much easier to detect and crystallize.
4. The hinge binders, which are the most common place for fragments to bind to kinases, provide a very nice potential for several molecular interactions. Specifically, 1–3 hydrogen bonds and a lipophilic pocket can all contribute to high ligand efficiency.
5. The commonality of ATP binding sites has allowed the ability to transfer X-ray binding mode information from one kinase to another. This greatly expands the number of scaffolds that can be utilized for fragment optimization.

## Acknowledgement

The author is indebted to Drs. M. Vieth, J. Sutherland, J. Toth, D. Robertson, and C. Humblet for valuable suggestions and support of this chapter.

## References

1. Fabbro D, Cowan-Jacob SW, Mobitz H, Martiny-Baron G (2012) Targeting cancer with small-molecular-weight kinase inhibitors. *Methods Mol Biol* 795:1–34
2. Arencibia JM, Pastor-Flores D, Bauer AF, Schulze JO, Biondi RM (2013) AGC protein kinases: from structural mechanism of regulation to allosteric drug development for the treatment of human diseases. *Biochim Biophys Acta* 1834:1302–1321
3. Cohen P, Alessi DR (2013) Kinase drug discovery – what's next in the field? *ACS Chem Biol* 8:96–104
4. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. *Science* 298(1912–1916):1933–1934
5. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
6. Breitenlechner CB, Bossemeyer D, Engh RA (2005) Crystallography for protein kinase drug design: PKA and SRC case studies. *Biochim Biophys Acta* 1754:38–49
7. Chiu Y-Y, Lin C-T, Huang J-W, Hsu K-C, Tseng J-H, You S-R, Yang J-M (2013) KIDFamMap: a database of kinase-inhibitor-disease family maps for kinase inhibitor selectivity and binding mechanisms. *Nucleic Acids Res* 41:D430–D440
8. Ghose AK, Herbertz T, Pippin DA, Salvino JM, Mallamo JP (2008) Knowledge based prediction of ligand binding modes and rational inhibitor design for kinase drug discovery. *J Med Chem* 51:5149–5171
9. Liao JJ-L (2007) Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J Med Chem* 50:409–424
10. van Linden OP, Kooistra AJ, Leurs R, de Esch IJ, de Graaf C (2014) KLIFS: a knowledge-based structural database to navigate kinase-ligand interaction space. *J Med Chem* 57(2):249–277
11. Vieth M, Sutherland JJ, Robertson DH, Campbell RM (2005) Kinomics: characterizing the therapeutically validated kinase space. *Drug Discov Today* 10:839–846
12. Bossemeyer D, Engh RA, Kinzel V, Ponstingl H, Huber R (1993) Phosphotransferase and

- substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 Å structure of the complex with manganese(2+) adenylyl imidodiphosphate and inhibitor peptide PKI(5–24). *EMBO J* 12:849–859
13. Wang Z, Canagarajah BJ, Boehm JC, Kassisa S, Cobb MH, Young PR, Abdel-Meguid S, Adams JL, Goldsmith EJ (1998) Structural basis of inhibitor selectivity in MAP kinases. *Structure (London)* 6:1117–1128
  14. Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach DR, Miller WT, Clarkson B, Kuriyan J (2002) Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (ST1-571). *Cancer Res* 62:4236–4243
  15. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274:1531–1534
  16. Chessari G, Woodhead AJ (2009) From fragment to clinical candidate – a historical perspective. *Drug Discov Today* 14:668–675
  17. Zartler ER, Shapiro MJ (2005) Fragonomics: fragment-based drug discovery. *Curr Opin Chem Biol* 9:366–370
  18. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46:3–26
  19. Congreve M, Carr R, Murray C, Jhoti H (2003) A ‘rule of three’ for fragment-based lead discovery? *Drug Discov Today* 8:876–877
  20. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* 9:430–431
  21. Vieth M, Siegel MG, Higgs RE, Watson IA, Robertson DH, Savin KA, Durst GL, Hipskind PA (2004) Characteristic physical properties and structural fragments of marketed oral drugs. *J Med Chem* 47:224–232
  22. Saxty G, Woodhead SJ, Berdini V, Davies TG, Verdonk ML, Wyatt PG, Boyle RG, Barford D, Downham R, Garrett MD, Carr RA (2007) Identification of inhibitors of protein kinase B using fragment-based lead discovery. *J Med Chem* 50:2293–2296
  23. Verdonk ML, Rees DC (2008) Group efficiency: a guideline for hits-to-leads chemistry. *ChemMedChem* 3:1179–1180
  24. Leeson PD, Springthorpe B (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* 6:881–890
  25. Murray CW, Blundell TL (2010) Structural biology in fragment-based drug design. *Curr Opin Struct Biol* 20:497–507
  26. Gradler U, Bomke J, Musil D, Dresing V, Lehmann M, Holzemann G, Greiner H, Esdar C, Krier M, Heinrich T (2013) Fragment-based discovery of focal adhesion kinase inhibitors. *Bioorg Med Chem Lett* 23:5401–5409
  27. Lepre CA (2011) Practical aspects of NMR-based fragment screening. *Methods Enzymol* 493:219–239
  28. Lepre CA, Moore JM, Peng JW (2004) Theory and applications of NMR-based screening in pharmaceutical research. *Chem Rev* 104:3641–3675
  29. Neumann T, Junker HD, Schmidt K, Sekul R (2007) SPR-based fragment screening: advantages and applications. *Curr Top Med Chem* 7:1630–1642
  30. Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat Rev Drug Discov* 9:23–27
  31. Allen KN, Bellamacina CR, Ding X, Jeffery CJ, Mattos C, Petsko GA, Ringe D (1996) An experimental approach to mapping the binding surfaces of crystalline proteins. *J Phys Chem* 100:2605–2611
  32. Jhoti H, Cleasby A, Verdonk M, Williams G (2007) Fragment-based screening using X-ray crystallography and NMR spectroscopy. *Curr Opin Chem Biol* 11:485–493
  33. Davies TG, Tickle IJ (2012) Fragment screening using X-ray crystallography. *Top Curr Chem* 317:33–59
  34. Spurlino JC (2011) Fragment screening purely with protein crystallography. *Methods Enzymol* 493:321–356
  35. Leach AR, Hann MM, Burrows JN, Griffen EJ (2006) Fragment screening: an introduction. *Mol BioSyst* 2:429–446
  36. Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H (2005) Fragment-based lead discovery using X-ray crystallography. *J Med Chem* 48:403–413
  37. Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. *J Med Chem* 39:2887–2893
  38. Fejzo J, Lepre CA, Peng JW, Bemis GW, Ajay Murcko MA, Moore JM (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem Biol* 6:755–769
  39. Baurin N, Aboul-Ela F, Barril X, Davis B, Drysdale M, Dymock B, Finch H, Fromont C, Richardson C, Simmonite H, Hubbard RE (2004) Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets. *J Chem Inf Comput Sci* 44:2157–2166

40. Aronov AM, McClain B, Stuver MC, Murcko MA (2008) Kinase-likeness and kinase-privileged fragments: toward virtual polypharmacology. *J Med Chem* 51:1214–1222
41. Muegge I, Enyedy IJ (2004) Virtual screening for kinase targets. *Curr Med Chem* 11:693–707
42. Vieth M, Erickson J, Wang J, Webster Y, Mader M, Higgs R, Watson I (2009) Kinase inhibitor data modeling and de novo inhibitor design with fragment approaches. *J Med Chem* 52: 6456–6466
43. Xia X, Maliski EG, Gallant P, Rogers D (2004) Classification of kinase inhibitors using a Bayesian model. *J Med Chem* 47:4463–4470
44. Sutherland JJ, Higgs RE, Watson I, Vieth M (2008) Chemical fragments as foundations for understanding target space and activity prediction. *J Med Chem* 51:2689–2700
45. Verdonk ML, Giangreco I, Hall RJ, Korb O, Mortenson PN, Murray CW (2011) Docking performance of fragments and druglike compounds. *J Med Chem* 54:5422–5431
46. Good AC, Liu J, Hirth B, Asmussen G, Xiang Y, Biemann H-P, Bishop KA, Fremgen T, Fitzgerald M, Gladysheva T, Jain A, Jancsics K, Metz M, Papoulis A, Skerlj R, Stepp JD, Wei RR (2012) Implications of promiscuous Pim-1 kinase fragment inhibitor hydrophobic interactions for fragment-based drug design. *J Med Chem* 55:2641–2648
47. Miranker A, Karplus M (1991) Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins Struct Funct Genet* 11:29–34
48. Clark M, Guarneri F, Shkurko I, Wiseman J (2006) Grand canonical Monte Carlo simulation of ligand-protein binding. *J Chem Inf Model* 46:231–242
49. Clark M, Meshkat S, Talbot GT, Carnevali P, Wiseman JS (2009) Fragment-based computation of binding free energies by systematic sampling. *J Chem Inf Model* 49:1901–1913
50. Bohm HJ (1992) The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J Comput Aided Mol Des* 6:61–78
51. Bohm HJ (1992) LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des* 6:593–606
52. Eisen MB, Wiley DC, Karplus M, Hubbard RE (1994) HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins Struct Funct Genet* 19:199–221
53. Lauri G, Bartlett PA (1994) CAVEAT: a program to facilitate the design of organic molecules. *J Comput Aided Mol Des* 8:51–66
54. Yang Y, Nesterenko DV, Trump RP, Yamaguchi K, Bartlett PA, Drueckhammer DG (2005) Virtual hydrocarbon and combinatorial databases for use with CAVEAT. *J Chem Inf Model* 45:1820–1823
55. Maass P, Schulz-Gasch T, Stahl M, Rarey M (2007) Recore: a fast and versatile method for scaffold hopping based on small molecule crystal structure conformations. *J Chem Inf Model* 47:390–399
56. Bohacek RS, McMurtin C (1994) Multiple highly diverse structures complementary to enzyme binding sites: results of extensive application of a de novo design method incorporating combinatorial growth. *J Am Chem Soc* 116:5560–5571
57. Todorov NP, Dean PM (1998) A branch-and-bound method for optimal atom-type assignment in de novo ligand design. *J Comput Aided Mol Des* 12:335–349
58. Todorov NP, Dean PM (1997) Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design. *J Comput Aided Mol Des* 11:175–192
59. Lewell XQ, Judd D, Watson S, Hann M (1998) RECAP-retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J Chem Inf Comput Sci* 38:511–522
60. Alberts IL, Todorov NP, Dean PM (2005) Receptor flexibility in de novo ligand design and docking. *J Med Chem* 48:6585–6596
61. Dean PM, Firth-Clark S, Harris W, Kirton SB, Todorov NP (2006) SkelGen: a general tool for structure-based de novo ligand design. *Expert Opin Drug Discov* 1:179–189
62. Stahl M, Todorov NP, James T, Mauser H, Boehm H-J, Dean PM (2002) A validation study on the practical use of automated de novo design. *J Comput Aided Mol Des* 16:459–478
63. Pierce AC, Rao G, Bemis GW (2004) BREED: generating novel inhibitors through hybridization of known ligands. Application to CDK2, P38, and HIV protease. *J Med Chem* 47:2768–2775
64. Schrödinger Release 2014-4: Maestro, version 10.0, (2014) Schrödinger, LLC, New York, NY
65. Molecular Operating Environment (MOE), 2013.08; (2015) Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7
66. Sutherland JJ, Nandigam RK, Erickson JA, Vieth M (2007) Lessons in molecular recognition. 2. Assessing and improving cross-docking accuracy. *J Chem Inf Model* 47:2293–2302
67. Nandigam RK, Evans DA, Erickson JA, Kim S, Sutherland JJ (2008) Predicting the accuracy

- of ligand overlay methods with random forest models. *J Chem Inf Model* 48:2386–2394
68. Erickson JA, Mader MM, Watson IA, Webster YW, Higgs RE, Bell MA, Vieth M (2010) Structure-guided expansion of kinase fragment libraries driven by support vector machine models. *Biochim Biophys Acta* 1804:642–652
  69. Erlanson DA, McDowell RS, O'Brien T (2004) Fragment-based drug discovery. *J Med Chem* 47:3463–3482
  70. deKloe GE, Bailey D, Leurs R, deEsch IJP (2009) Transforming fragments into candidates: small becomes big in medicinal chemistry. *Drug Discov Today* 14:630–646
  71. Wyatt PG, Woodhead AJ, Berdini V, Boulstridge JA, Carr MG, Cross DM, Davis DJ, Devine LA, Early TR, Feltell RE, Lewis EJ, McMenamin RL, Navarro EF, O'Brien MA, O'Reilly M, Reule M, Saxty G, Seavers LCA, Smith D-M, Squires MS, Trewartha G, Walker MT, Woolford AJA (2008) Identification of N-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *J Med Chem* 51:4986–4999
  72. Howard S, Berdini V, Boulstridge JA, Carr MG, Cross DM, Curry J, Devine LA, Early TR, Fazal L, Gill AL, Heathcote M, Maman S, Matthews JE, McMenamin RL, Navarro EF, O'Brien MA, O'Reilly M, Rees DC, Reule M, Tisi D, Williams G, Vinkovic M, Wyatt PG (2009) Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multitargeted kinase inhibitor with potent aurora kinase activity. *J Med Chem* 52:379–388
  73. Antonysamy S, Hirst G, Park F, Sprengeler P, Stappenbeck F, Steensma R, Wilson M, Wong M (2009) Fragment-based discovery of JAK-2 inhibitors. *Bioorg Med Chem Lett* 19: 279–282
  74. Xiang Y, Hirth B, Asmussen G, Biemann H-P, Bishop KA, Good A, Fitzgerald M, Gladysheva T, Jain A, Jancsics K, Liu J, Metz M, Papoulis A, Skerlj R, Stepp JD, Wei RR (2011) The discovery of novel benzofuran-2-carboxylic acids as potent Pim-1 inhibitors. *Bioorg Med Chem Lett* 21:3050–3056
  75. Chu MLH, Lang Z, Chavas LMG, Neres J, Fedorova OS, Tabernero L, Cherry M, Williams DH, Douglas KT, Evers PA (2010) Biophysical and X-ray crystallographic analysis of Mps1 kinase inhibitor complexes. *Biochemistry* 49:1689–1701
  76. Matthews TP, Klair S, Burns S, Boxall K, Cherry M, Fisher M, Westwood IM, Walton MI, McHardy T, Cheung K-MJ, Van MR, Williams D, Aherne GW, Garrett MD, Reader J, Collins I (2009) Identification of inhibitors of checkpoint kinase 1 through template screening. *J Med Chem* 52:4810–4819
  77. Reader JC, Matthews TP, Klair S, Cheung K-MJ, Scanlon J, Proisy N, Addison G, Ellard J, Piton N, Taylor S, Cherry M, Fisher M, Boxall K, Burns S, Walton MI, Westwood IM, Hayes A, Eve P, Valenti M, Brandon AH, Box G, Montfort RLM, Williams DH, Aherne GW, Raynaud FI, Eccles SA, Garrett MD, Collins I (2011) Structure-guided evolution of potent and selective CHK1 inhibitors through scaffold morphing. *J Med Chem* 54:8328–8342
  78. Lainchbury M, Matthews TP, McHardy T, Boxall KJ, Walton MI, Eve PD, Hayes A, Valenti MR, de HBAK, Box G, Aherne GW, Reader JC, Raynaud FI, Eccles SA, Garrett MD, Collins I (2012) Discovery of 3-alkoxyamino-5-(pyridin-2-ylamino)pyrazine-2-carbonitriles as selective, orally bioavailable CHK1 inhibitors. *J Med Chem* 55: 10229–10240
  79. Li R, Martin MP, Liu Y, Wang B, Patel RA, Zhu J-Y, Sun N, Pireddu R, Lawrence NJ, Li J, Haura EB, Sung S-S, Guida WC, Schonbrunn E, Sebti SM (2012) Fragment-based and structure-guided discovery and optimization of rho kinase inhibitors. *J Med Chem* 55: 2474–2478
  80. Irie T, Fujii I, Sawa M (2012) Design and combinatorial synthesis of a novel kinase-focused library using click chemistry-based fragment assembly. *Bioorg Med Chem Lett* 22:591–596
  81. Gentile G, Bernasconi G, Pozzan A, Merlo G, Marzorati P, Bamborough P, Bax B, Bridges A, Brough C, Carter P, Cutler G, Neu M, Takada M (2011) Identification of 2-(4-pyridyl)thienopyridinones as GSK-3 $\beta$  inhibitors. *Bioorg Med Chem Lett* 21:4823–4827
  82. Erlanson DA, Arndt JW, Cancilla MT, Cao K, Elling RA, English N, Friedman J, Hansen SK, Hession C, Joseph I, Kumaravel G, Lee W-C, Lind KE, McDowell RS, Miatkowski K, Nguyen C, Nguyen TB, Park S, Pathan N, Penny DM, Romanowski MJ, Scott D, Silvian L, Simmons RL, Tangonan BT, Yang W, Sun L (2011) Discovery of a potent and highly selective PDK1 inhibitor via fragment-based drug discovery. *Bioorg Med Chem Lett* 21:3078–3083
  83. Tsai J, Lee JT, Wang W, Zhang J, Cho H, Mamo S, Bremer R, Gillette S, Kong J, Haass NK, Sproesser K, Li L, Smalley KSM, Fong D, Zhu Y-L, Marimuthu A, Nguyen H, Lam B, Liu J, Cheung I, Rice J, Suzuki Y, Luu C, Settachatgul C, Shellooe R, Cantwell J, Kim

- S-H, Schlessinger J, Zhang KYJ, West BL, Powell B, Habets G, Zhang C, Ibrahim PN, Hirth P, Artis DR, Herlyn M, Bollag G (2008) Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc Natl Acad Sci U S A* 105: 3041–3046
84. Flaherty KT, Yasothan U, Kirkpatrick P (2011) Vemurafenib. *Nat Rev Drug Discov* 10: 811–812

# Chapter 14

## Designing a Small Molecule Erythropoietin Mimetic

Frank Guarnieri

### Abstract

Erythropoietin (EPO) is a protein made by the kidneys in response to low red blood cell count that is secreted into the bloodstream and binds to a receptor on hematopoietic stem cells in the bone marrow inducing them to become new red blood cells. EPO made with recombinant DNA technology was brought to market in the 1980s to treat anemia caused by kidney disease and cancer chemotherapy. Because EPO infusion was able to replace blood transfusions in many cases, it rapidly became a multibillion dollar per year drug and as the first biologic created with recombinant technology it launched the biotech industry. For many years intense research was focused on creating a small molecule orally available EPO mimetic. The Robert Wood Johnson (RWJ) group seemed to definitively establish that only large peptides with a minimum of 60 residues could replace EPO, as anything less was not a full agonist. An intense study of the published work led me to hypothesize that the size of the mimetic is not the real issue, but the symmetry making and breaking of the EPO receptor induced by the ligand is the key to activating the stem cells. This analysis meant that residues in the binding site of the receptor deemed absolutely essential for ligand binding and activation from mutagenesis experiments, were probably not really that important. My fundamental hypotheses were: (a) the symmetric state of the homodimeric receptor is the most stable state and thus must be the off-state, (b) a highly localized binding site exists at a pivot point where the two halves of the receptor meet, (c) small molecules can be created that have high potency for this site that will be competitive with EPO and thus can displace the protein–protein interaction, (d) small symmetric molecules will stabilize the symmetric off-state of the receptor, and (e) a key asymmetry in the small molecule will stabilize a mirror image asymmetry in the receptor resulting in the stabilization of the on-state and proliferation of the stem cells into red blood cells. Researchers at Amgen published a co-crystal structure of EPO bound to the EPO receptor, which has a beautiful twofold symmetry—it was argued that this is the active state of the receptor. Activating the EPO receptor with EPO induces an almost instantaneous shutdown mechanism to sharply curtail any proliferative signal transduction, and thus, my hypotheses lead to the conclusion that the Amgen co-crystal is actually the state after receptor downregulation and thus an off-state. To put these hypotheses to the test, my computational method of Simulated Annealing of Chemical Potential was run using the co-crystal created at RWJ, which is the receptor trapped in a partial agonist state. The simulations predicted a previously unknown high affinity binding site at the pivot point where the two halves of the dimeric receptor meet, and detailed analysis of the fragment patterns led to the prediction of a molecule less than 300 MW that is basically twofold symmetric with a chiral center on one side and not the other. Thus, to the degree that computer simulations can be taken seriously, these results support my hypotheses on small molecule receptor activation. When this small molecule was synthesized and tested it indeed induced human hematopoietic stems cells to become red blood cells. When the predicted chiral center of this molecule was removed eliminating its one asymmetric feature, the resulting molecule was an antagonist—it could potently displace hot EPO but could no longer induce stem cell proliferation and differentiation. These results provided strong support for my theories on how to create potent small molecule EPO agonists and

were used to launch the new company Locus Pharmaceuticals. These molecules, however, required significant chemical changes in order to make them stable in other in vitro assays and to be in vivo active, but these alterations had to be done in a way that maintained the symmetry–asymmetry considerations that led to the creation of an in vitro active molecule. The combination of changing functional groups to enable good pharmacokinetics, while not changing the key intrinsic symmetry properties were never seriously pursued at Locus and the program died. Investigations into how red blood cells are created have occupied many prominent researchers for the entire twentieth century. In the second half of the century EPO was discovered and by the end of the century it became a blockbuster commercial product that launched the biotech revolution.

**Key words** Erythropoietin, EPO mimetic, Small molecule, Protein–protein interaction, Fragment based drug design, Monte Carlo

---

## 1 Historical Background

What are the factors that induce the bone marrow to produce new red blood cells? The search for answers to this question occupied many investigators during the twentieth century and the findings have the dual distinctions of being an epic case study in medical research and a focal point of the genomics revolution. In the first half of the 1900s it was generally accepted that low blood oxygen levels directly stimulated the bone marrow to produce new red blood cells. Thus, it is interesting that as early as 1906 Carnot [1, 2] proposed the alternate theory of a humoral factor produced by some organ in response to low oxygen levels that circulated to the bone marrow stimulating red blood cell production. For the next 40 years, no research group produced compelling experimental evidence supporting either theory. In 1950, Reissmann [3] parabiotically united a pair of rats by an anastomosis from the ears to the tail roots and demonstrated that elevated erythropoiesis occurred in both partners when only one was subjected to hypoxic conditions. This was arguably the first significant experimental demonstration of the existence of a blood factor, the so-called erythropoietin (EPO), responsible for causing the bone marrow to produce red blood cells. The interpretation of Reissmann's results is that the animal put in the hypoxic chamber produces excess EPO, while its partner in the normal atmospheric chamber does not. The normal animal nevertheless, has enhanced levels of red blood cells that are comparable to its hypoxic partner, because the EPO produced by one is shared by the other through the anastomosis.

Although Riessmann's experiments were a tour-de-force, the complexity and expertise required to carry out the anastomosis procedure made replication of results by other labs difficult to impossible. For the community to definitively adopt the EPO theory of red blood cell production, simple widely accessible procedures that could be reproduced by others needed to be developed. In 1953 Erslev [4] showed that normal rabbits would produce large amounts of reticulocytes 4–6 days after being injected with

200 ml of plasma drawn from anemic rabbits but there were no measurable changes in reticulocyte count if the injected plasma was taken from normal animals. The interpretation of these results is that rabbits made anemic either by starvation or bleeding produce EPO in order to stimulate the bone marrow to create new replacement red blood cells. Since the plasma collected from anemic rabbits should be high in EPO—if this “anemic” plasma is injected into normal rabbits these animals should produce significantly enhanced numbers of reticulocytes. Erslev’s procedure is simple and easy to reproduce by other labs. The experimental procedures became solidified in 1955 when Jacobson and coworkers [5] demonstrated that the EPO effect could be quantitated by monitoring the red blood cell uptake of radioactive iron-59. By injecting the test animals with radiolabeled iron in conjunction with anemic plasma and then subsequently harvesting the cellular fraction from blood, an assay of the radioactivity gives quantitative information on how EPO stimulates red blood cell production.

Once the EPO theory of red cell proliferation became embraced by the medical community, the next step was to find the organ that produced it. Jacobson and coworkers [6] made their next contribution to this field. They systematically removed different organs in rats and then subjected them to either a significant hemorrhage or a dose of CoCl<sub>2</sub>. Only nephrectomy resulted in no elevation of EPO. The interpretation of these results is that hypophysectomy or thyroidectomy or splenectomy or removal of most of the liver followed by hemorrhage or CoCl<sub>2</sub> red cell destruction, did not stop the rats from producing large amounts of EPO—plasma redrawn from these animals could dramatically increase red cell production in normal rodents. When the kidneys were removed, however, no EPO production was induced in the animals. These results were taken as definitive proof that the kidneys are the source of EPO. Subsequent studies by Reissmann et al. [7] and Hirashima and Takaku [8] confirmed that the kidneys are in fact the source of EPO, unleashing an explosive amount of new investigations and necessitating that researchers have readily available sources of EPO and a generally agreed upon reference standard. In 1966 the World Health Organization certified an International Reference Preparation of Erythropoietin [9] isolated from severely anemic human urine making 1 ml samples available to the research community with an arbitrary assignment of 10 units of EPO activity.

Continuing research throughout the 1970s clearly indicated that EPO would very likely be an effective treatment for various forms of anemia [10] especially in renoprival [11] patients. The detailed molecular characterization necessary for further progress required highly purified protein, which Goldwasser and Kung produced [12] in 1971, but unfortunately in quantities too small for sequence analysis. It took another 6 years for these investigators [13] to develop a protocol for producing purified EPO in milligram quantities.

This was of seminal importance in developing EPO as a commercial product, because the Miyake purification method was used to produce protein for tryptic digestion, sequence analysis of peptide fragments, DNA probe construction, genomic library screening [14], and cloning and expression of the EPO gene by both Amgen [15] and Genetics Institute [16]. Just 2 years later in 1987, recombinant human erythropoietin from Amgen was shown in clinical trials [17] to restore hematocrit to normal levels in patients with end-stage renal failure with no discernible toxicities, thus eliminating the need for transfusions in these patients. For more than two decades to the present recombinant EPO has been used for the treatment of cancer chemotherapy induced anemia and renal failure, with both [18] having 2003 sales ~\$4B for Procrit from Johnson and Johnson and Epogen and Aranesp [19] combined from Amgen, respectively.

---

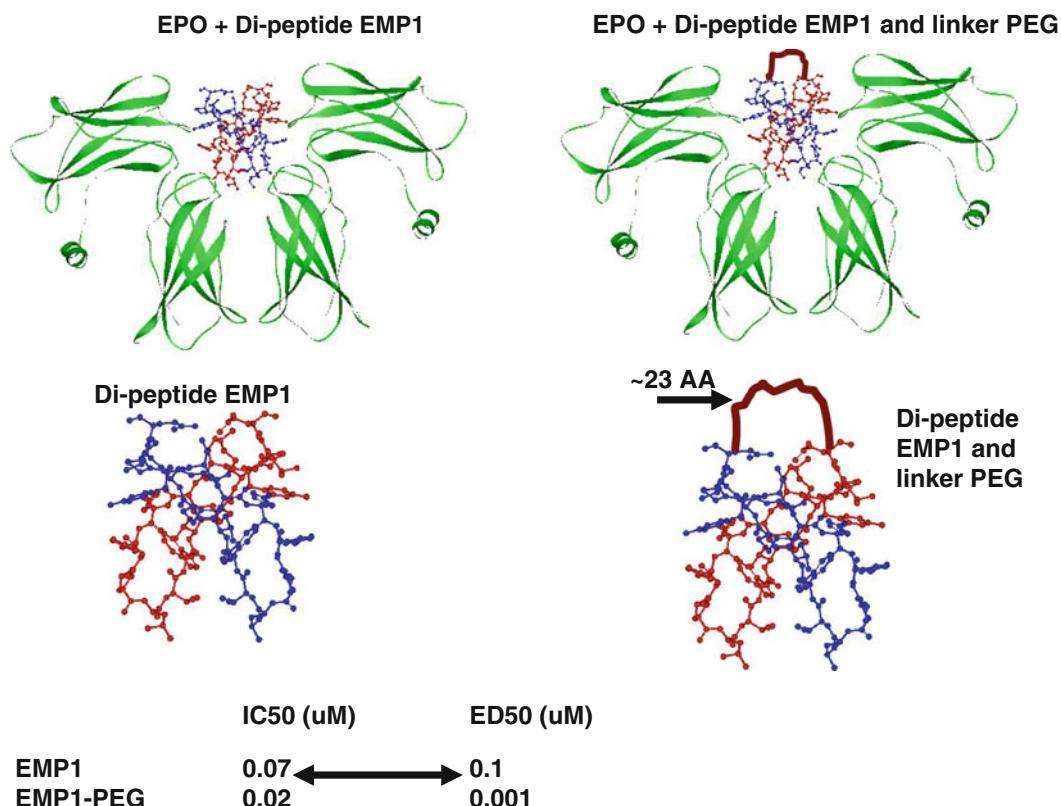
## 2 The Case Against the Existence of an Orally Available Small Molecule EPO Mimetic

The earliest evidence suggesting that any EPO mimetic whatsoever might be impossible is posttranslational modification of the protein. Both recombinant studies cited above [15, 16] state that about 40 % of EPO's molecular weight comes from carbohydrates. Physicochemical comparisons between natural EPO purified from the urine of patients with aplastic anemia and recombinant EPO indicate that both [20] have one O-glycosylation and three N-glycosylation sites. Wasley and coworkers [21] have shown that EPO with incompletely processed N-linked oligosaccharides is 500-fold less active in vivo. Experiments from Yamaguchi and colleagues [22] confirm this finding while Delorme et al. produced [23] similar results by selectively mutating sites of glycosylation. One reasonable interpretation of these results is that a subset of the posttranslational modifications are essential for high in vivo EPO activity. In fact, Amgen's Aranesp is a modified non-natural version of EPO specifically engineered [24] to have five N-glycosylation sites, which gives it a threefold longer serum half-life, thus further validating the importance of posttranslational modification.

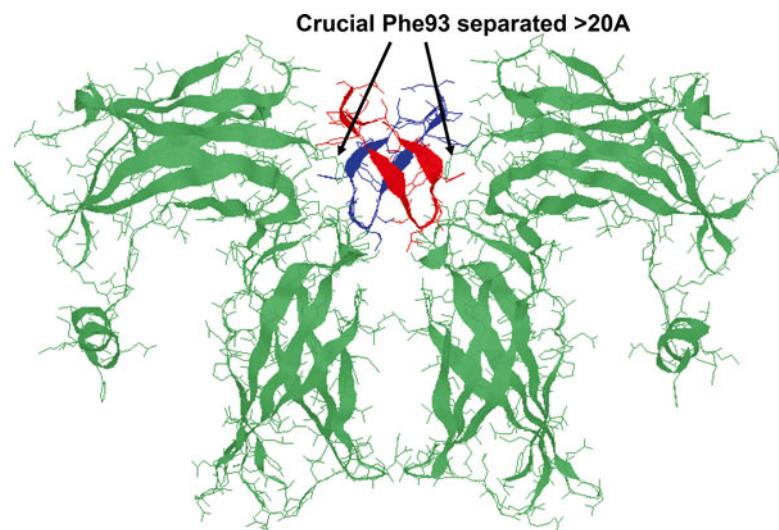
### 2.1 The Work from Robert Wood Johnson (RWJ)

The RWJ group produced a singularly impressive body of work that is the basis of the atomic level mechanistic understanding of the binding and signal transduction of EPO. RWJ in collaboration with Affymax [25] discovered 20-mer EPO mimetic peptides (EMP) using phage display and showed that EMPI is an in vitro and in vivo agonist. RWJ solved the co-crystal structure [26] of the extracellular binding domain of the EPO receptor (EPOR) with EMPI demonstrating homodimerization of EPOR with two equivalents of bound ligand. The EMPI dimer is approximately two orders of

magnitude less potent than EPO, so the RWJ group had the clever idea of cross-linking the EMP1–EMP1 dimer, hypothesizing that this should increase agonist potency. They discovered that a polyethylene glycol linker [27] that is approximately 23 amino acids long yielded a molecule EMP1–PEG–EMP1 that has comparable agonist potency to EPO itself. These results, summarized in Fig. 1, definitively demonstrate that EPO mimetics that are functional both *in vitro* and *in vivo* can be created and that the apparently essential posttranslational modification of the natural hormone is in fact not an absolute requirement for activating the receptor. The other strong conclusion of the RWJ work is that ~60 AA is the minimal structure for full agonist mimetic activity (EPO is 166 AA with almost 50 % of its total weight coming from carbohydrates)



**Fig. 1** A summary of some results from the Robert Wood Johnson Group. The *upper left panel* is the co-crystal structure of the 20-mer EPO mimetic peptide known as EMP1 (Science 1996). Both the receptor and ligand crystallize in the dimeric state. The *upper right hand panel* is a simplified rendition of cross-linking of EMP1 with polyethyleneglycol (Chem and Bio 1997) also known as PEG. The number of repeating units in the linker is approximately 23 amino acids long. The *lower left and right hand panels* are a blowup of the EMP1 dimer and EMP1 cross-linked with PEG. Interestingly, shown at the *bottom*, with or without cross-linking the binding affinity is almost the same as seen in the IC50, but agonist activity is enhanced by two orders of magnitude with the addition of the PEG cross-link shown in the ED50



**Fig. 2** The RWJ group identified the pair of Phe93 residues as critical interaction points for both natural EPO and EMP1. The salient point that they make is that EMP1 has no sequence similarity to EPO and yet both make intimate interactions with the pair of Phe93 amino acids as demonstrated by mutagenesis experiments (JBC 1997). The interpretation of these results is the PEG-EMP1 is likely the minimal ligand that can have full agonist activity, because this construct is required for proper ligand presentation to the pair of Phe93 residues of the receptor

with additional supporting evidence coming from their mutagenesis studies showing that ligand contact with Phe93 on both receptors [28] is essential for receptor activation. The pair of Phe93 residues is separated by over 20A as shown in Fig. 2. Taken as a whole, this collection of work demonstrates that carbohydrates are not necessarily needed for ligand agonist activity, but that there is no way to make a small organic molecule with agonist potency comparable to EPO—at least not one that is small enough to be potentially orally bioavailable if dual Phe93 interaction is required.

## 2.2 Non-peptide Organic EPO Mimetics

In 1999 the Merck group [29] did in fact demonstrate that organic non-peptide molecules could bind to and activate the EPOR. Their strategy was to first identify a small organic molecule that binds to the receptor by screening an in-house library and then covalently link multiple copies of this molecule to a common core. The screening hit (*N*-3-[2-(4-biphenyl)-6-chloro-5-methyl]indolyl-acetyl-L-lysine methyl ester) inhibited EPO binding with an IC<sub>50</sub> of 59.5 μM. Eight copies of this molecule attached to an eight-fold reactive polyamidoamino-octa-4-hydroxymethylbenzamide dendrimer resulted in a molecule that inhibited EPO binding with an IC<sub>50</sub> of 4.4 μM. The dendrimer scaffold alone had no detectable binding to the EPOR. This very complicated large small

molecule “octa-oligomer” does in fact show weak agonist activity in stimulating human hematopoietic progenitor cells, which does establish that non-peptides can activate the EPOR. A group at Scripps took a similar approach to the Merck group, creating and screening combinatorial libraries designed to mimic the RWJ EMP1 peptide to obtain weak EPOR binders and then oligomerizing these small molecules unto symmetric cores to enhance potency and demonstrate weak agonist activity [30]. While these efforts demonstrate that organic structures can activate the EPOR, the size and complexity of these molecules seem to indicate the impossibility of making a small drug-like EPO agonist.

---

### 3 Theory and Hypotheses for Creating Small Molecule EPO Mimetics

The RWJ group performed mutagenesis studies [31] confirming that the pair of Phe93 residues are important binding partners for both EPO itself and the mimetic peptides, which led them to conclude that “these residues may represent a minimum epitope on the EPOR for productive ligand binding.” Figure 2 shows the distance between the pair of Phe93 and the space below that needs to be filled in, in order to create a scaffold to present the interactions to these key residues, which can only lead to the conclusion that no small molecule can be created to activate the receptor. The Merck and Scripps work discussed above apparently confirms these conclusions, because their organic molecules were designed to mimic these key interactions. A small molecule in this context means something that has the potential to be orally bioavailable—a compound that has a molecular weight of less than 600. The combination of all the work summarized so far indicates that this is not possible.

The Amgen group solved the crystal structure of EPO bound to the extracellular binding domain. It is a beautiful twofold symmetric ligand–receptor complex, which “shows that erythropoietin imposes a unique 120° angular relationship and orientation that is responsible for optimal signaling through intracellular kinase pathways [32].” The orientation of the receptor in this EPO bound structure is significantly different from the structure solved by the RWJ group with bound EPO mimetic peptides. The EPO mimetic peptides are partial agonists, so they do in fact activate the receptor to some extent even though the receptor is in a completely different state compared to the Amgen structure. My interpretation of Amgen’s results is quite different.

#### 3.1 Hypothesis I: The Amgen Structure Is the Receptor in the Off-State Not the On-State

There are two main reasons for my hypothesis, one biological and one mathematical. The biological reason is that EPOR normally resides in the off state, because it is immediately shut down by intramolecular phosphorylation and ubiquitination [33] upon activation. Any growth factor receptor has to normally be in the

off-state because being excessively in the on-state as is the case with constitutively active mutants [34] can lead to cancer. From this it may be concluded that the on-state is a higher energy state and that it is probably unstable requiring ligand-binding to populate it. The straightforward interpretation of the EPO-bound co-crystal of the receptor is that this is the ligand induced active state. The act of forming crystals that yield high resolution structure, however, inherently drives the system into a stable state *not a transient active state*; thus, I hypothesize that the Amgen structure represents an off-state. Mathematically, the most stable state of a complex system is often the state of highest symmetry, the global minimum, which I hypothesize, is represented by this structure.

**3.2 Hypothesis II:  
The Partial Agonist  
State of the EPOR  
Represented by  
the EMP1 Bound  
Peptides from RWJ  
Has an Asymmetry  
That Is Induced by  
the Ligand**

**3.3 Hypothesis III:  
Interacting  
with the Essential  
Phe93 Pair Is Not  
Essential for Receptor  
Activation**

**3.4 Hypothesis IV:  
The Pivot Point Where  
the Two Receptors  
Meet Is a High Affinity  
Binding Site  
of the Receptor  
and a Small Molecule  
That Binds at This Site  
Will Activate  
the Receptor if It Has  
the Right Symmetry  
Properties**

This is essentially a corollary to the first hypothesis. If the off-state of the receptor is the stable 2-fold symmetric state, then activating the receptor requires that a ligand break this symmetry.

The RWJ group has shown very compelling experimental evidence on the seminal importance of interacting with the Phe93 pair of the dimerized EPO receptor for both EPO and EPO mimetic peptide agonist activity. They also showed that inducing receptor dimerization alone is not sufficient [35] for activation. The essence of this hypothesis is that receptor activation occurs via a ligand-driven switch from the symmetric off-state to its asymmetric on-state. It should be possible to induce this change without interacting with the Phe93 pair.

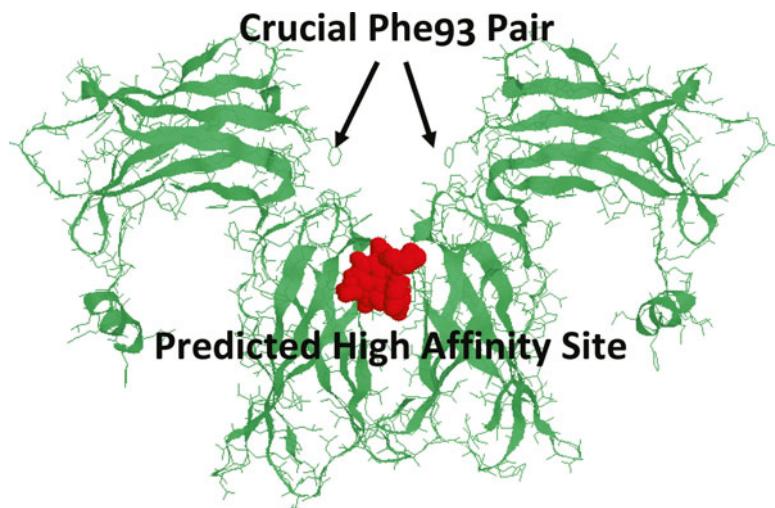
The essence of this last hypothesis is on the nature of the dimerized receptors, particularly the point where they meet and how the two parts pivot either symmetrically or asymmetrically relative to each other. This pivoting will cause the receptor to be in the off-state or the on-state, respectively. A small molecule that binds to this site must have an asymmetry that counteracts the asymmetry of the activated receptor. When they are paired, the asymmetry of one cancels the asymmetry of the other resulting in a symmetric ligand-receptor complex that transiently stabilizes the active state of the receptor.

## 4 Testing the Hypotheses with Computer Simulations

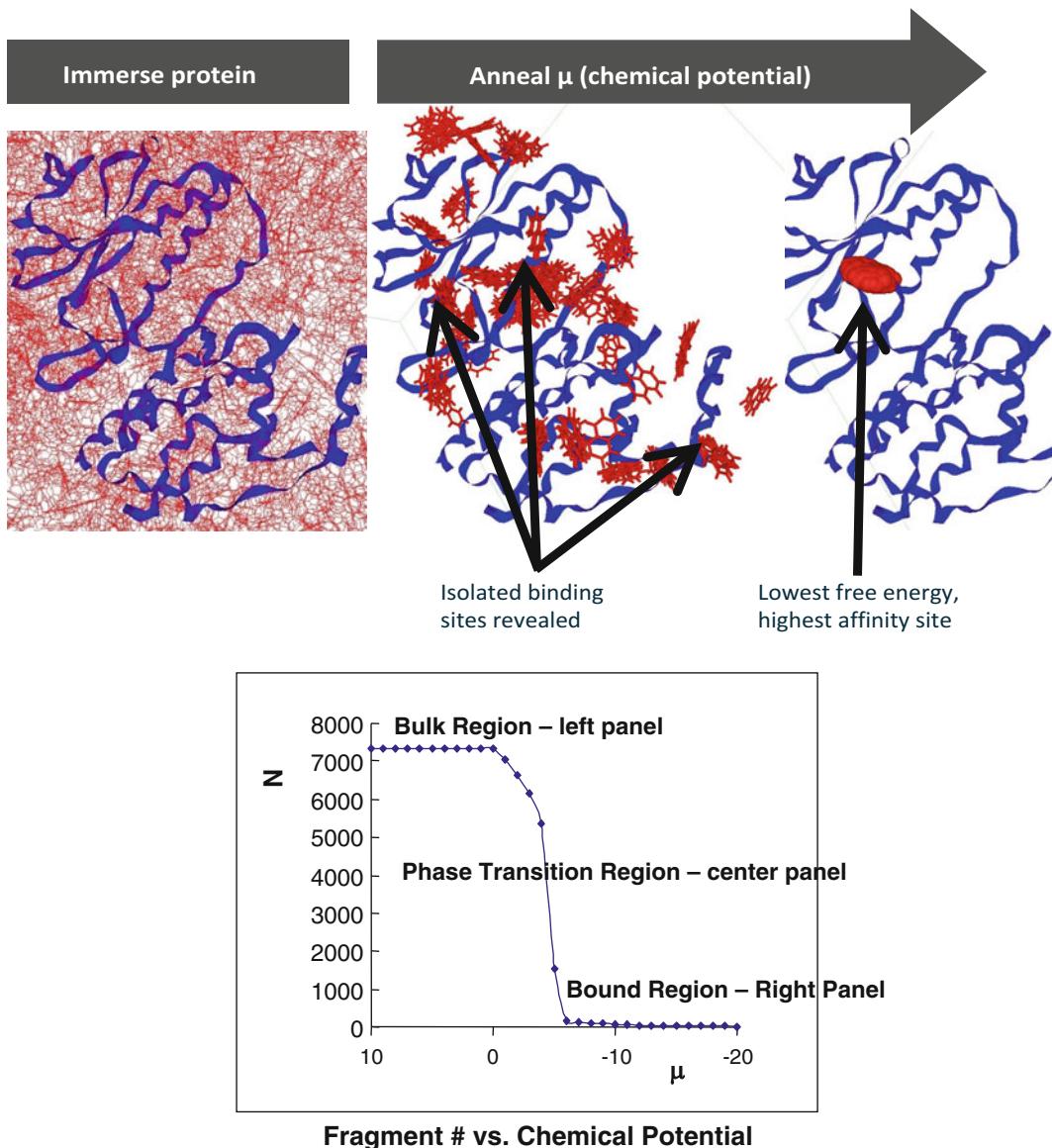
To my knowledge, there is no experimental evidence that the pivot point where the two receptors meet has any importance whatsoever for ligand binding and receptor activation. So the first prerequisite is to test the hypothesis that this location is a high affinity binding site—and this will be done with simulations. The computational technique of grand canonical Monte Carlo [36] with the variable chemical potential [37] method (SACP) that I first introduced to study the differential hydration properties of the major and minor grooves of DNA [38] has been generalized to predict protein binding sites. The algorithm successfully predicted the binding sites in eight different variants of hen egg white lysozyme [39], demonstrating its robustness, and predicted a previously unknown lipid binding site [40] in mitochondrial aspartate amino transferase (mAAT). Experiments on the mAAT confirmed the computational predictions that this enzyme binds lipids and moonlights [40] as a plasma membrane transporter. The SACP simulations do indeed predict that a high affinity binding site exists at the point where the two EPO receptors meet (Fig. 3).

### 4.1 Simulated Annealing of Chemical Potential: The Method

SACP is a straightforward Monte Carlo procedure that is illustrated in Fig. 4. A simplified but representative protocol is to generate a random number ( $\text{RanX}$ ) between 0 and 1 and to insert



**Fig. 3** The red blob is a cluster of high affinity fragments predicted by the computer simulation technique simulated annealing of chemical potential (SACP). In order to be classified as a true binding site this location must also have a predicted low affinity for water. So at least by computer simulation, there is evidence that a high affinity binding site does exist where the two receptors meet in support of Hypothesis IV. Note SACP has been very successful in predicting binding sites



**Fig. 4** The SACP method is demonstrated schematically. The graph at the bottom shows that when the chemical potential of the protein simulation cell is high, the bulk region at  $B=10$ , the probability of accepting an insertion of a fragment is very likely and the simulation cell becomes completely saturated, which is shown in the upper left panel. As the chemical potential is lowered, nothing particularly dramatic occurs, because insertions and deletions are approximately balanced. When the chemical potential is lowered to the point that it begins to surpass the free energy of fragment–fragment cohesion (this is the solvent–solvent interaction energy), the system goes through a phase transition evacuating most of the solvent fragment molecules, which is shown in the center panel. The very interesting result is that a discrete number of sites on the protein retain fragment binding throughout the phase transition—these are the high affinity sites. As the chemical potential is annealed (further lowered) additional sites are evacuated, which is shown in the right panel. Because chemical potential is a formal free energy, this process defines a quantitative relative rank order affinity between different sites on the protein for a particular fragment. Different fragments have different binding patterns

a fragment into the protein simulation cell if  $\text{RanX} < 0.5$  or delete a fragment if  $\text{RanX} > 0.5$ . In either case this is a trial move with attempted insertion assigned a probability of  $P = \exp[-[E(n+1) - E(n)]/(RT)] \times \exp[B]/(N+1)$  and attempted fragment deletion a probability of  $P = N \times \exp[-[E(n-1) - E(n)]/(RT)] \times \exp[-B]$  where  $E(n)$  is the energy of the system before an insert or delete,  $E(n+1)$  is the energy of the system after a fragment is inserted, and  $E(n-1)$  is the energy of the system after a fragment has been deleted,  $R$  is the gas constant and  $T$  is the temperature in Kelvin.  $B$  is the chemical potential of the simulation cell containing the protein and the only adjustable parameter. Operationally, a random number ( $\text{RanX}$ ) is generated between 0 and 1 and the inserted or deleted fragment is accepted as the new configuration if  $P \leq \text{RanX}$ . When  $B$  is set high, the probability of inserting a fragment into the protein simulation cell is dramatically enhanced resulting in saturation with fragments as shown in the left panel of Fig. 4. As the chemical potential is gradually lowered by decreasing the value of  $B$ , fragments persist in the simulation cell until a low enough  $B$ -value occurs that causes a dramatic evacuation of the simulation cell with almost all fragments exiting. As shown in the middle panel of Fig. 4, discrete high affinity binding sites for one particular fragment are revealed. The chemical potential can be successively lowered until only one site remains as shown in the right-most panel of Fig. 4. Generally about five million simulation steps are done at each  $B$ -value (value of the chemical potential) and the simulation is run for about 30 different  $B$ -values. The delta- $B$ -value between two different positions on the protein for a given fragment is a measure of the relative free energy of binding since chemical potential is a formally correct free energy and not just an enthalpy of binding—sampling in this manner takes the entropy component into account. Note, when using somewhat more sophisticated procedures such as cavity-bias [41] to determine where to insert fragments more efficiently, the probability equation has to be adjusted to maintain detailed balance, but this is a technical matter that does not change the conceptual process. It is very important to note that the only input into the algorithm is the protein structure and the fragments to be simulated. There is no other human intervention so the predictions of high affinity vs. low affinity sites and quantitative rank ordering of binding is predicted in a totally objective manner. Flaws in the predictions occur when the force field parameters do not accurately represent the fields of the fragments or proteins, which is especially true of large concentrated charges. We have recently demonstrated these problems [42] in a blind competition study.

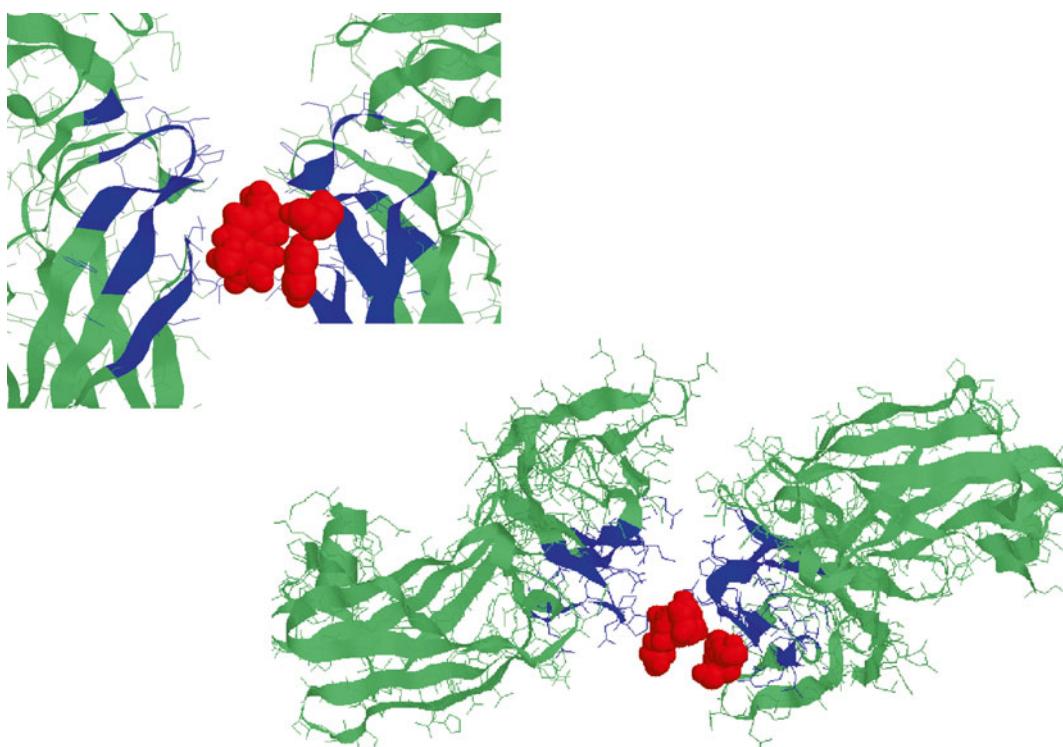
#### **4.2 Binding Site Determination**

One very compelling aspect of SACP is how the procedure can be generalized to create an automated protocol for locating sites on a protein capable of making high affinity non-bonded interactions,

so-called hotspots. When SACP is run on a range of fragments independently, each fragment has a distinct and unique binding pattern on the protein that is quite different from any other fragment. What is very striking is what happens when individual simulations are brought together and two questions are asked, (1) which localized sites on the protein have high affinity for a chemical diversity of fragments and thus can be labeled sites of clustering, and (2) of these clustering sites, which do NOT have high affinity for water (the so-called water exclusion principle). The combination of these three principles, SACP predicting high affinity fragment binding, clustering, and water exclusion have been shown to accurately predict protein hotspots in a wide variety of macromolecules.

#### **4.3 Prediction of Ligand Binding Asymmetry: Hypotheses II, III, and IV**

Figure 5 is a blowup of the predicted binding site of the EPOR shown in Fig. 3 from two different perspectives. These views clearly show not only that the simulation predicts a high affinity binding site at the point where the receptors meet but that the collective fragment patterns bind in an asymmetric mode. The locus of



**Fig. 5** Two different perspectives of the predicted high affinity binding site illustrated in Fig. 3 are shown so that the asymmetry of the fragment patterns can be clearly seen. Since SACP produces results that are independent of human intervention other than the choice of protein structure and the fragments to be simulated, this at least provides computational evidence for the perhaps surprising hypotheses that not only is the focal point where the receptors meet a high affinity binding site, but that it will have a characteristic asymmetry

predicted fragment binding from SACP is very small, which is an indication from the simulation of the potential for designing very small molecules with significant binding affinity. It is important to note that the Phe93 pair is far from this predicted binding site. This does not mean that SACP predicts that the Phe93 pair is unimportant. Examination of Fig. 4 shows that there are multiple important interaction sites on a protein. For the EPOR, SACP is predicting that the site of primary importance is the focal point where the receptors meet.

#### 4.4 What SACP Does

SACP takes a protein and a collection of organic fragments including water and runs a sequence of grand canonical ensemble simulations starting from a very high chemical potential with gradual reduction until the chemical potential is very low. Each fragment is a solvent molecule and is run on one CPU. If 100 fragments including water are run, then this can be carried out on 100 processors simultaneously. Each fragment is driven through a phase transition represented by the graph shown in Fig. 4. This usually produces about a dozen high affinity sites on a protein for a given fragment. As the chemical potential is lowered further, high affinity sites are gradually depopulated of fragments—this gives a relative rank order of the free energy of interaction between different protein localities for a given fragment. Because this performs a complete sampling of the protein–fragment interactions for the entire protein with no human bias and no a prior knowledge, it is an extremely powerful algorithm.

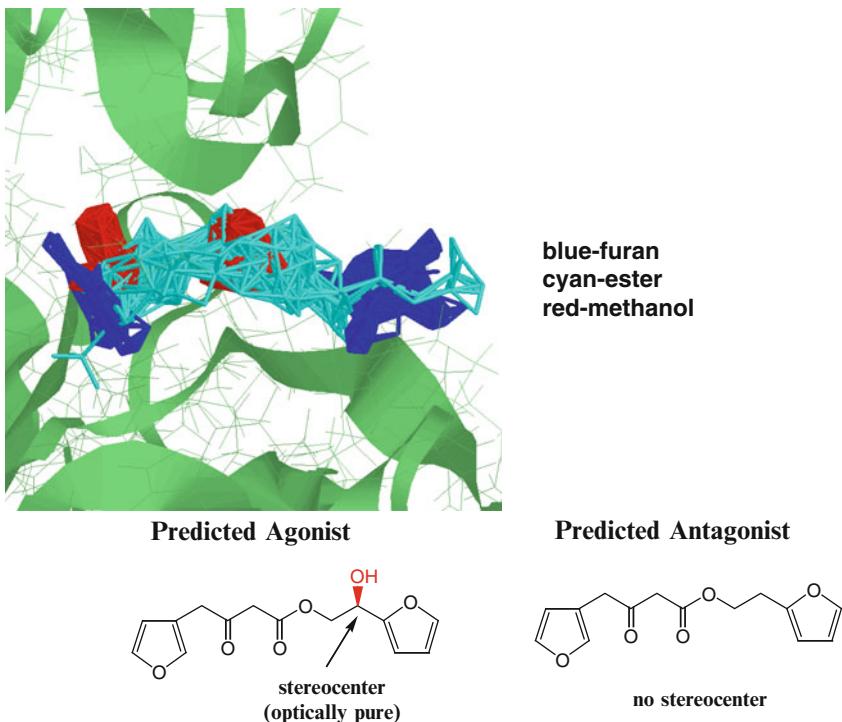
#### 4.5 What SACP Does NOT Do

The SACP method on its own will not determine or predict binding sites. When the prediction of fragment binding affinity as a function of chemical potential is augmented with the clustering and water exclusion postulates, then a very robust recipe for binding site prediction exists. SACP cannot possibly predict what will be an agonist for the EPO receptor or if such a thing even exists. The four hypotheses stated above were developed from a careful study of erythropoiesis, which is completely independent of SACP. Given these hypotheses, the fragment patterns generated from SACP can be analyzed to see if the simulation data supports or negates the preconceived notions. The structure of the SACP method does NOT allow for any biases to be built into the simulation, so the fragment patterns are unbiased. There is, however, a large bias that occurs with the choice of protein structure, because it is kept static during the simulations. My hypotheses lead to the perhaps counterintuitive choice of NOT using the Amgen structure with bound EPO, but the RWJ structure with the artificial EMP1 mimetic peptides. SACP has nothing to say about which protein structure should be used. Of course making an agonist is much more difficult than making an

antagonist—so designing an activator requires more a prior hypotheses and thus a deeper understanding of the system, whereas often making an antagonist requires imposing no hypotheses. To be clear, it is my personal choice to use the RWJ structure with the artificial EPO mimetic peptides removed in order to generate fragment patterns that I believe will lead to an agonist. I believe that using the Amgen structure co-crystallized with the natural EPO hormone will produce symmetric fragment patterns that will lead to the design of an antagonist. After choice of input structure is made, SACP will produce free energy rank-ordered fragment binding modes.

#### **4.6 Computational Design of the Proposed EPOR Agonist**

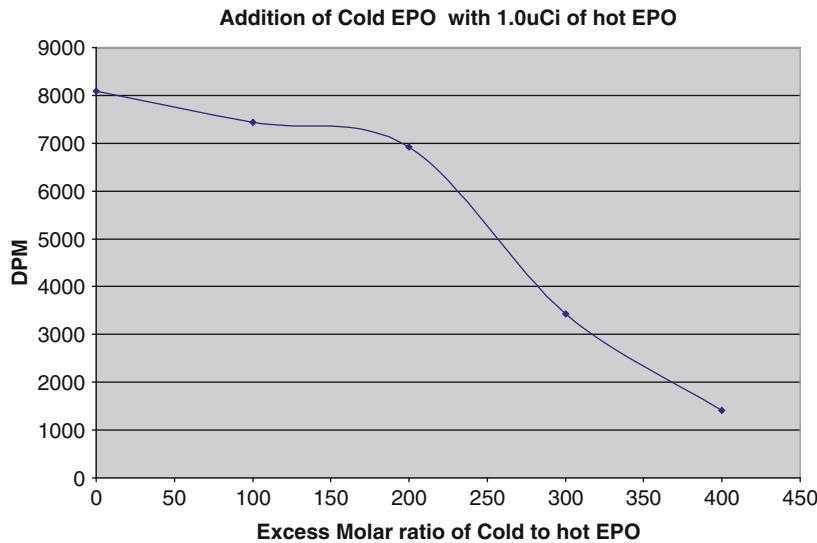
The actual molecular design process requires adding a little more detail to the hypotheses. Because the receptor in all states (on-state, off-state, or partial agonist state) has a basic twofold symmetry, I expect that the asymmetry of the on-state will be subtle. Chemically, asymmetry means a chiral center. Therefore, I expect that the fragment patterns will be basically symmetric from the SACP simulations, *but a chiral center will be predicted on one side and not the other*. The top panel of Fig. 6 shows a blowup of the EPOR binding site with overlaid high affinity fragment patterns of furan, methyl acetate, and methanol predicted by SACP. Furan shown in blue has a directional and distinct binding mode with a definite orientation that is twofold symmetric across both halves of the receptor dimer. Methylacetate populates the region between the furan pair with a translational invariance that merges into and links the furan duo with multiple copies, but shows distinct asymmetric preference for binding the carboxyl groups to one side. The equipotent multi-copy translational binding forms the di-ketone and ester moiety. The most striking pattern is the prediction of methanol. One methanol fragment cluster occurs in the dead center of the receptor dimer in an isolated fashion unable to bind with the other fragments. The other methanol cluster binds to only one side of the receptor and not the other. Furthermore, this methanol cluster merges into the methylacetate linker in a way that produces a chiral center. The data is only computational at this point, but SACP predicts, (1) a previously unknown high affinity binding site at the focal point where the receptors meet, (2) fragments that bind in a twofold symmetric manner with high affinity (very negative chemical potential)—the furan pair, (3) introduction of a stereocenter on one side of the molecule and not the other and thus a characteristic asymmetry. *The bottom left panel of Fig. 6 shows the molecule that I designed from these patterns, which my hypotheses predict will be an agonist.* The bottom right panel shows the molecule with the stereocenter removed, and thus, my hypotheses predict that this should be an antagonist.



**Fig. 6** The *top panel* shows a blowup of three different superimposed high affinity fragment patterns predicted by SACP in the previously unknown binding site of the EPO-R using the structure from the RWJ group. The input structure for the simulations was the receptor bound to the EMP1 peptides with the EMP1 peptides removed. Furan, shown in *blue*, binds in a twofold symmetric manner with a clearly defined orientation. Methylacetate has a set of translationally invariant equipotent high affinity states that merge into a linker for the furan pair. Methylacetate does NOT bind in a twofold symmetric state according to the SACP simulations, but places the ester group preferentially to one side creating a double ketone moiety. One methanol cluster binds right in the center of where the two receptors meet. What is most striking is the high affinity binding of methanol that merges into one side of the acetate linker but is completely absent from the other. At least the computational data support the hypotheses that an asymmetry exists in the receptor that will give rise to a complementary asymmetry in the ligand. It also suggests that if the molecule has the stereocenter removed, it will no longer be able to activate the receptor and thus should be an antagonist, which is shown in the *lower right panel*

## 5 Binding and Proliferation Experimental Results

Theories and hypotheses are thought exercises that make science fun and interesting and computer simulations while intriguing may or may not capture an important aspect of reality. Ultimately, experiments must be performed to test the predictions, so the molecules shown in Fig. 6 were synthesized and tested in binding and proliferation experiments. A total of three molecules were made, both stereoisomers of the optically active molecule and the molecule without the stereocenter. It is important to note that only one stereoisomer is definitely favored according to the SACP simulations, indicating that only one enantiomer will be an agonist.



**Fig. 7** EPO binding experiments are commonly done by competing radioactive I-125 labeled EPO against cold EPO. Hot EPO is purchased in a highly pure form and is quite expensive. Cold EPO is a dilute mixture and thus is significantly cheaper, so C-EPO is used wherever possible. The molar excess of C-EPO needed to displace H-EPO must be established, which is 400 $\times$

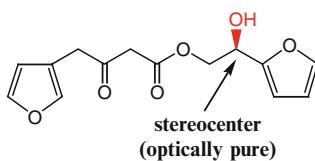
Binding experiments are commonly standardized by first establishing competition between I-125-EPO (hot or H-EPO) and cold EPO (C-EPO). Hot EPO is purchased highly purified and cold EPO is purchased as a dilute mixture. We performed these set of experiments multiple times and determined that displacement of H-EPO required a 400-fold molar excess of C-EPO, which is shown in Fig. 7 with the experimental protocol shown in Fig. 8. While the experimental protocol needs to be followed carefully, it is conceptually straightforward. Erythroid progenitor cells are proliferated with EPO demonstrating that functional receptor exists. The cells are then washed and centrifuged to remove any EPO. Equivalent amounts of these cells are added to a set of centrifuge tubes. C-EPO in 400-fold molar excess relative to H-EPO is pipetted into a subset of control tubes. Various concentrations of the predicted small molecule agonist and antagonist are pipetted into select tubes that are labeled. All tubes then get H-EPO in the concentration precisely matching what is displaced by the 400X C-EPO. The cells are incubated for 3 h then washed to remove any unbound H-EPO. Each tube is individually measured with a counter to quantitate the remaining bound H-EPO. These experiments were replicated many times with representative results shown in Fig. 9. Both compounds have high affinity for the EPOR with the predicted agonist being about 10 nm and the predicted antagonist being about 100 nm. *It is quite amazing and satisfying that molecules with molecular weights of less than 280 can displace H-EPO with such potency.*

### Binding Procedure

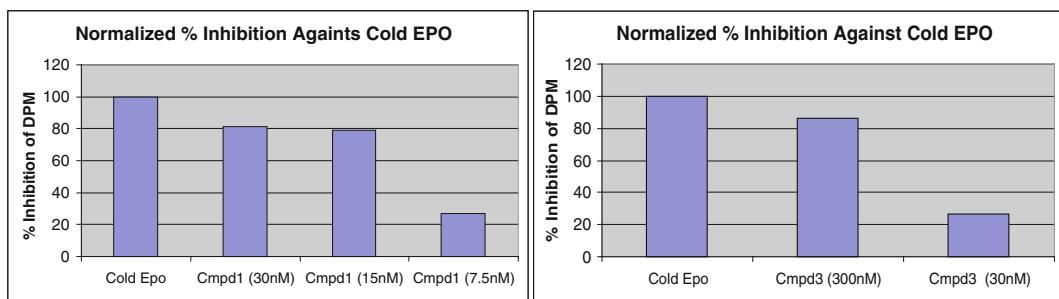
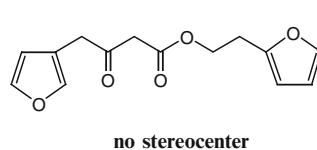
- Grew Erythroid Progenitor cells in DPBM Medium containing IL-3, IL-6, SCF and EPO
- Spun the cells down for 10 minutes at 1,000rpm
- Added 3M Glycine pH 3 in 0.9% Saline to the cells for 3 minutes to dissociate preformed receptor-ligand complexes
- Washed the cells with DPBM medium containing 5% BSA
- Counted and resuspended the cells in final volume of  $1 \times 10^6$  cell/200ul
- Aliquoted 200ul of cells in each centrifuge tube
- Let the cells sit on ice for ~5minutes
- Added 400X cold Epo to the control tubes
- Added compounds (30nM-7.5nM) to the labeled tubes
- Added PBS/0.1% BSA to all the tubes except cold Epo to compensate for addition of cold Epo
- Added PBS/0.1% BSA/0.00042% DMSO to the all the tubes except compound tubes to compensate for addition of compounds
- Diluted  $I^{125}$ Epo in DPBM Medium/2-5% BSA bringing final concentration to 0.5uci/100ul
- Added 100ul of  $I^{125}$ Epo solution to all the tubes
- Incubated the the tubes for 3 hours on the rotator at 4°C
- Took 10ul aliquots from each tube for counting the cells
- Added 800ul of Dibutyl Phthalate to all the tubes
- Shook the tubes and centrifuged them for 2 minutes at 10,000g
- Took supernatant off , froze the bottom of the tubes in Liquid Nitrogen, and clipped the bottoms into LSC vials
- Added 100ul of PBS buffer to the LSC vials to dissolve the pellet
- Added 5ml of Ready Safe to the vials, shook them and counted in LSC6500

**Fig. 8** This is the protocol for a standard assay used to quantitate EPO and EPO mimetic binding. Cells need to be proliferated with EPO to make sure that they express functional receptor. The EPO then needs to be washed out. The same number of cells is added to all centrifuge tubes. C-EPO in 400× molar excess relative to H-EPO is added to control tubes in order to prevent all H-EPO binding. A set of tubes get different concentrations of compounds. H-EPO is added to all tubes and the tubes are incubated for 3 h. All tubes are washed to remove unbound H-EPO. Each tube is put into a scintillation counter to quantitate bound H-EPO

Predicted Agonist



Predicted Antagonist



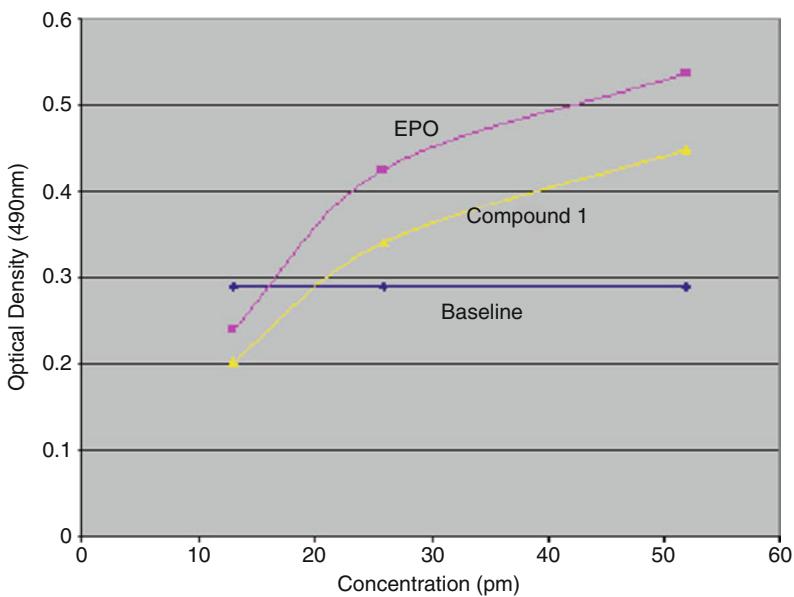
**Fig. 9** The compounds do indeed bind to the EPO receptor and displace H-EPO. The compound with the stereocenter has a molecular weight of only 277 Da and yet has an IC50 of about 10 nm. Removing the hydroxyl functional group that creates the stereocenter reduces the binding to about 100 nm

## Proliferation Procedure

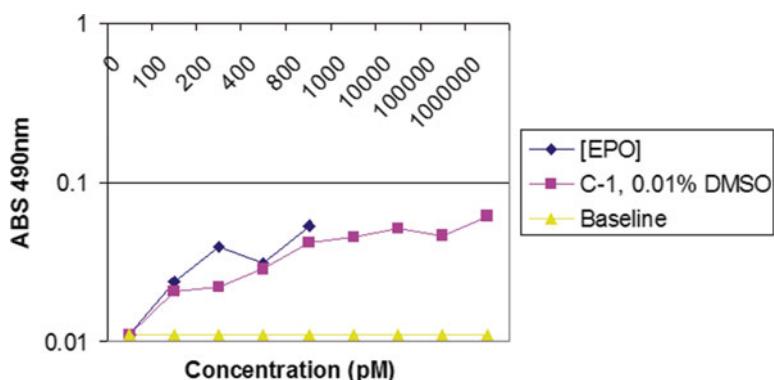
- Spun down cells (1000 rpm/15min)
- Removed supernatant
- Resuspended pellet in 30 ml of Epo-free media (DPBM media, SCF, IL-3, and IL-6)
- Spun down cells (700 rpm/10 min)
- Removed supernatant
- Resuspended cells in 30 ml of Epo-free media and starved cells for 2hrs at 37°C/5%CO<sub>2</sub>
- Spun down cells (700 rpm/10 min)
- Counted cells via hemacytometer grid method
- To each flask added 10 ml of Epo-free media and 500,000 cells
- Added appropriate amount of Epo or Test compound to flask
- Grew cells at 37°C/5% CO<sub>2</sub>
- MTS Assay
  - Aliquoted out 100ul/200ul sample of cell suspension into 96-well plate
  - (each flask is sampled 5X)
  - Added 20ul of Solution One
  - Incubated plates for 3 – 4 hrs at 37°C/5% CO<sub>2</sub>
  - Read plate on Dynex plate reader at 490nm (reference filter 630nm)
- Hand Count Assay
  - Spun down 5 ml of cell suspension (1000 rpm/3 min)
  - Removed 4.5 ml of supernatant
  - Resuspended cells
  - To a 50 ul aliquot and added 50 ul of trypan blue stain
  - Loaded aliquot onto hemacytometer and counted cells

**Fig. 10** Standard experimental protocols were used to determine cell proliferation. The MTS One Solution assay uses a tetrazolium salt that is converted to a colored soluble Formazan by an enzyme in living cells, and thus, the absorption intensity is directly proportional to the number of cells. The older method used MTT, which is converted into a crystalline Formazan product that needs to be isolated and dissolved in order to perform colorimetric assays. All results were confirmed by a technician using a hemocytometer. This is a tedious procedure using a specialized microscope slide with an indented rectangular well that is precisely gridded on the millimeter and submillimeter scale. It enables depositing exact volumes of a solution containing cells into small wells in a way that a human being can look under the microscope and actually count the cells. This gives precise cell densities. It is common to use hemocytometers to confirm cell proliferation data

Testing whether or not these compounds can induce stem cells to divide and then become red blood cells is the final set of experiments. For this purpose we used human hematopoietic stem cells purchased from a blood bank. The standard MTS proliferation assay was used and all results were confirmed by cell counting with a hemocytometer. The proliferation experimental procedures are summarized in Fig. 10. The basic principle is very simple—enzymes in living cells will convert a tetrazolium salt into a colored formazan product and the intensity of this color is proportional to the number of living cells. The One Solution is a commercially available kit (Promega) for doing this assay. It is recommended to read the intensity at 490 nm and subtract out the

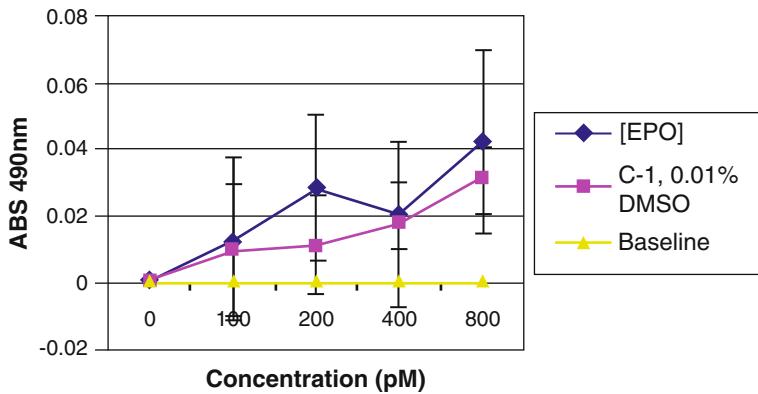


**Fig. 11** Human hematopoietic stem cells were used for the MTS proliferation assays. The optical density measurements show that the compound predicted to be an agonist causes cellular proliferation almost as potently as EPO itself

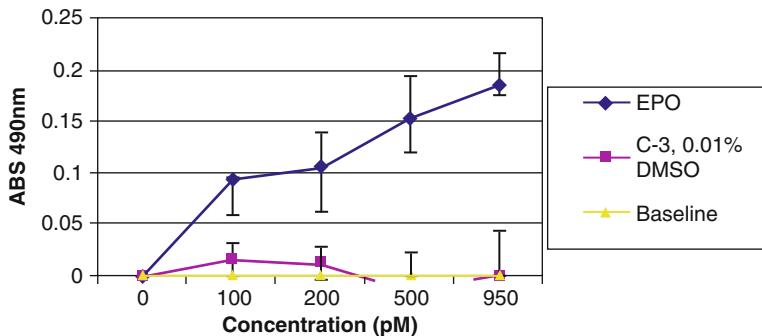


**Fig. 12** The assays were set up on a different day and redone—the results again show that the compound predicted to be an agonist causes cell proliferation comparable to EPO

absorbance at 620 or 630 nm. Figure 11 shows the results of our first experiment, which demonstrates that the predicted agonist compound is almost as potent as EPO. Note that the *X*-axis is in picomolar concentration units. The compound's binding affinity at 10 nm is reminiscent of the RWJ findings (Fig. 1) that receptor activation occurs at concentrations substantially below the IC<sub>50</sub>. This is indeed a very exciting result that strongly confirms the proposed hypotheses, so the assays were set up on another day and repeated. Figure 12 shows that the second time the assays were



**Fig. 13** The results were repeated in triplicate

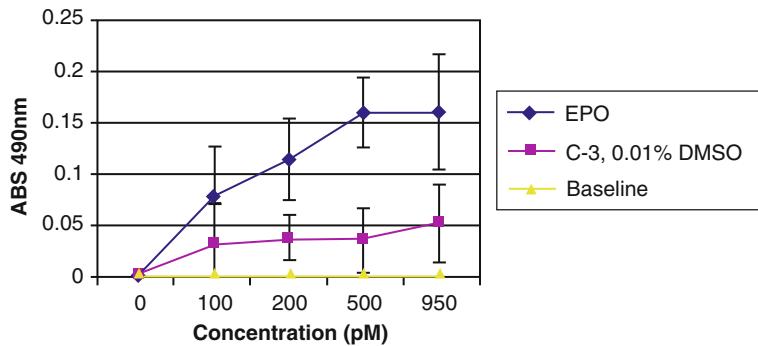


**Fig. 14** The results were done in triplicate, but this time using the compound with the stereocenter removed, showing that the compound cannot cause cellular proliferation

performed—again the compound is almost as potent as EPO. On yet another day the assays were set up in triplicate, which is shown in Fig. 13. All results were confirmed by cell counting with a hemocytometer. Next, the compound with the stereocenter removed was tested in the cellular proliferation assays in triplicate on 2 different days. These experiments show that this compound cannot cause proliferation of human hematopoietic stem cells as shown in Figs. 14 and 15.

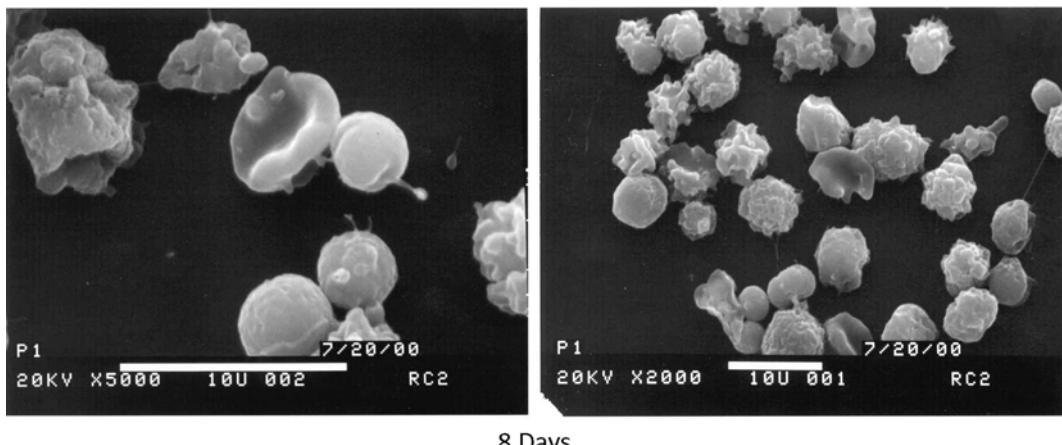
## 6 Scanning Electron Microscopy

Samples of the proliferation experiments were taken on different days and analyzed with scanning electron microscopy. The first SEM experiments were repeatedly performed on samples of human



**Fig. 15** The results were repeated

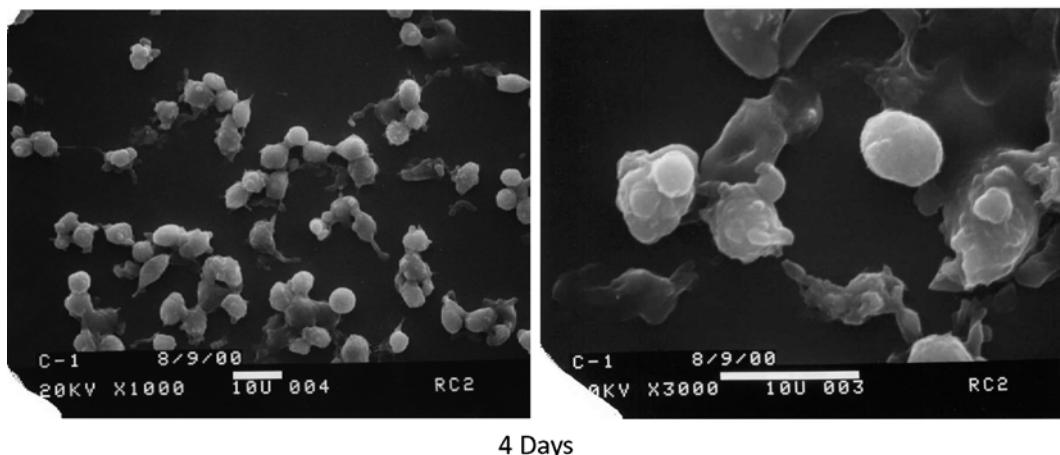
### SEM of EPO Induced Cells (IL-3, IL-6, SCF, EPO)



**Fig. 16** Scanning electron microscopy of EPO induced differentiation of hematopoietic stem cells into red blood cells. EM scans were done every day for 10 days—on the eighth day we obtained a picture of a puckered cell shown in the *left panel* that is characteristic of a red blood cell. The larger view in the *right panel* shows a few puckered cells—many others appear to be dead or dying

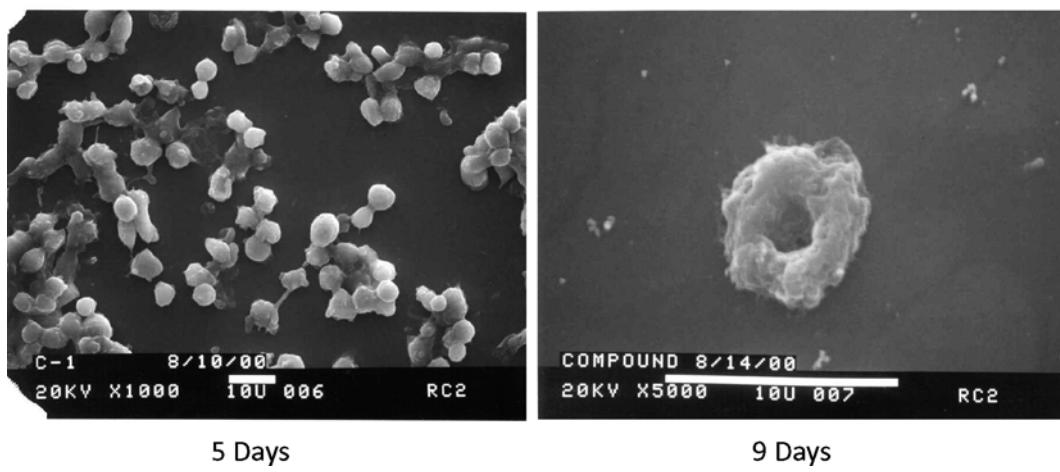
hematopoietic stem cells exposed to EPO. In our hands it took 8 days before observing the characteristic puckered red blood cell shown Fig. 16. SEM showed the compound inducing cell division on the fourth day as shown in the left panel of Fig. 17 with a puckered red blood cell shown in the right panel. An image on the fifth day shown in the left panel of Fig. 18 shows cell in the process of dividing as seen by the cell-cell contact apparently indicating one cell splitting off from the other and what look like several puckered red blood cells. An image from the ninth day shows one red blood cell that is apparently in very poor condition.

## SEM of C-1 Induced Cells (IL-3, IL-6, SCF, C-1)



**Fig. 17** After the fourth day, the EM scans show that the compound is inducing cell division as can be seen in the *left panel*. In the *right panel* is a puckered cell indicating that a red blood cell has been formed

## SEM of C-1 Induced Cells (IL-3, IL-6, SCF, C-1)



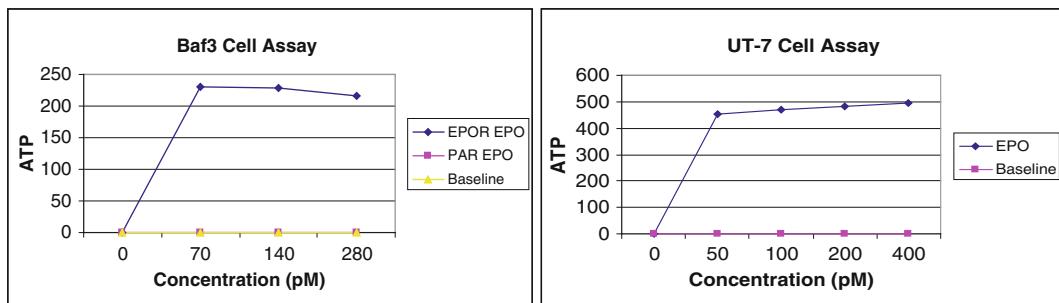
**Fig. 18** On the fifth day that hematopoietic stem cells are exposed to the compound one of the EM scans shows what appears to be many red blood cells in the *left panel*. On the ninth day only one red blood cell is left and is apparently in a very poor state

## 7 Conclusions

The results presented in this chapter are very exciting, because they lend strong support to a set of hypotheses on how the EPO receptor is activated and illustrate how Simulated Annealing of Chemical Potential can be used to design extremely low molecular weight novel compounds to test such hypotheses. Furthermore, the

fact that these compounds have a MW < 300 indicates that there is a possibility of making orally bioavailable small molecule EPO mimetic agonists. It must, however, be stated that the agonist compound is certainly not a drug and that several complex issues remain unresolved. The first most obvious flaw with these small molecules is the ester bond. Carboxylesterase [43] is the main enzyme in first pass hydrolysis of drugs and is the reason that most prodrug strategies [44] incorporate an ester bond, which will be cleaved in the intestine. The small molecules used in these studies here have a highly exposed ester bond and thus will be highly susceptible to hydrolytic cleavage. Prodrug studies have long recognized that special care is required to chemically stabilize [45] esters so further in vivo investigations on the small molecules created for these studies will likely require replacing or at least chemically stabilizing the ester bond. The most obvious next step of these investigations is solving the co-crystal structure of the extracellular domain of EPOR with the small molecules—the predicted binding mode needs to be confirmed. Secondly, SACP simulations should be applied to the Amgen structure, see if the fragment patterns confirm my hypothesis that these patterns will be symmetric and not asymmetric as occurred with the RWJ structure and that the resultant small molecules will be antagonists not agonists.

Finally, after these studies were concluded, a small amount of assay work was performed using Baf3 and UT-7 cell lines, which are commonly employed when working with EPO. Baf3 are an immortalized cell line derived from the bone marrow of rodents that requires IL-3 to survive and grow. When Baf3 cells are transfected with the EPOR gene, these cells can survive and grow when exposed to EPO, so this has become a common cell line in EPO research. There are, however, complications, because Krosl and coworkers [46] have shown signaling through the IL-3 pathway actually inhibits EPO activity in Baf3 cells. UT-7 is a cell line from leukemia patients that expresses EPOR and thus has become a standard tool in EPO research, because proliferation occurs upon application of EPO. The Centocor group [47] has done an intriguing study comparing a set of high potency EPO agonist biological molecules for their ability to cause proliferation in UT-7 cells versus in vivo activity. When EPO, Aranesp, and Centocor's antibody that contains two EMP1 peptides were compared, EPO and Aranesp caused UT-7 cellular proliferation 100x more potently than the Centocor construct—it is important to note that all constructs potently proliferated UT-7 cells. Conversely, the Centocor construct caused a sixfold greater increase in red blood cell count relative to EPO. These observations led the Centocor investigators to conclude that maintaining agonist activity over the time required for the maturation of reticulocytes is likely an important factor for in vivo efficacy. This would seem to disfavor an agonist with a highly exposed ester bond, which is expected to have a short half-life in vivo.



**Fig. 19** Baf3 are immortalized murine cells derived from bone marrow that depend upon IL3—withdrawing IL3 will cause the cells to go into apoptosis. Baf3 cells transfected with EPOR survive and proliferate upon application of EPO. UT-7 cells come from the bone marrow of patients with acute megakaryoblastic leukemia and proliferate when exposed to EPO. EPO activates signal transduction in both of these cell lines resulting in elevation of intracellular levels of ATP. EPO has no effect on ATP levels of Baf3 cells not transfected with EPOR. The predicted agonist compound has no effect on ATP levels in either of these cell lines

Since application of EPO to UT-7 or Baf3 cells transfected with EPOR activates signal transduction pathways, it was expected that this cellular activation would result in a rise of intracellular ATP levels. This is easily checked by applying EPO to the cells, lysing the cells, and measuring the ATP content [48] with a luciferase assay. EPO does indeed cause a significant rise in ATP of both cell lines as shown in Fig. 19. The agonist compound does NOT cause elevation of ATP levels in these cells, possibly because the ester bond is cleaved under the conditions of this assay—but this is unknown at this time.

## References

1. Carnot P, Deflandre C (1906) The hemopoietic activity of serum during the regeneration of blood. *C R Acad Sci* 143:384–387
2. Editorial (1972) Erythropoietin. *Br Med J* 1: 263
3. Reissmann KR (1950) Studies on the mechanism of erythropoietic stimulation in parabiotic rats during hypoxia. *Blood* 5:372–380
4. Erslev A (1953) Humoral regulation of red cell production. *Blood* 8:349–357
5. Plzak LF, Fried W, Jacobson LO, Bethard WF (1955) Demonstration of stimulation of erythropoiesis by plasma from anemic rats using Fe59. *J Lab Clin Med* 46:671–678
6. Jacobson LO, Goldwasser E, Fried W, Plzak L (1957) Role of the kidney in erythropoiesis. *Nature* 179:633–634
7. Reissmann KR, Nomura T, Gunn RW, Brosius F (1960) Erythropoietic response to anemia or erythropoietin injection in uremic rats with or without functioning renal tissue. *Blood* 16: 1411–1423
8. Hirashima K, Takaku F (1962) Experimental studies on erythropoietin. II. The relationship between juxtaglomerular cells and erythropoietin. *Blood* 20:1–8
9. Cotes PM, Bangham DR (1966) The international reference preparation of erythropoietin. *Bull World Health Organ* 35:751–760
10. Ward HP, Kurnick JE, Pisarczyk MJ (1971) Serum level of erythropoietin in anemias associated with chronic infection, malignancy, and primary hematopoietic disease. *J Clin Invest* 50:332–335
11. Radtke HW, Claussner A, Erbes PM, Scheuermann EH, Schoeppe W, Koch KM (1979) Serum erythropoietin concentration in chronic renal failure: relationship to degree of anemia and excretory renal function. *Blood* 54: 877–884

12. Goldwasser E, Kung CK (1971) Purification of erythropoietin. *Proc Natl Acad Sci U S A* 68: 697–698
13. Miyake T, Kung CK, Goldwasser E (1977) Purification of human erythropoietin. *J Biol Chem* 252:5558–5564
14. Lawn RM, Fritsch EF, Parker RC, Blake G, Maniatis T (1978) The isolation and characterization of linked delta- and beta-globin genes from a cloned library of human DNA. *Cell* 15:1157–1174
15. Lin FK, Suggs S, Lin CH, Browne JK, Smalling R, Egrie JC, Chen KK, Fox GM, Martin F, Stabinsky Z et al (1985) Cloning and expression of the human erythropoietin gene. *Proc Natl Acad Sci U S A* 82:7580–7584
16. Jacobs K, Shoemaker C, Rudersdorf R, Neill SD, Kaufman RJ, Mufson A, Seehra J, Jones SS, Hewick R, Fritsch EF et al (1985) Isolation and characterization of genomic and cDNA clones of human erythropoietin. *Nature* 313: 806–810
17. Eschbach JW, Egrie JC, Downing MR, Browne JK, Adamson JW (1987) Correction of the anemia of end-stage renal disease with recombinant human erythropoietin. Results of a combined phase I and II clinical trial. *N Engl J Med* 316:73–78
18. Pavlou AK, Reichert JM (2004) Recombinant protein therapeutics – success rates, market trends and values to 2010. *Nat Biotechnol* 22:1513–1519
19. Locatelli F, Vecchio LD (2001) Darbepoetin alfa. Amgen. *Curr Opin Investig Drugs* 2: 1097–1104
20. Imai N, Kawamura A, Higuchi M, Oh-edo M, Orita T, Kawaguchi T, Ochi N (1990) Physicochemical and biological comparison of recombinant human erythropoietin with human urinary erythropoietin. *J Biochem* 107: 352–359
21. Wasley LC, Timony G, Murtha P, Stoudemire J, Dorner AJ, Caro J, Krieger M, Kaufman RJ (1991) The importance of N- and O-linked oligosaccharides for the biosynthesis and in vitro and in vivo biologic activities of erythropoietin. *Blood* 77:2624–2632
22. Yamaguchi K, Akai K, Kawanishi G, Ueda M, Masuda S, Sasaki R (1991) Effects of site-directed removal of N-glycosylation sites in human erythropoietin on its production and biological properties. *J Biol Chem* 266: 20434–20439
23. Delorme E, Lorenzini T, Giffin J, Martin F, Jacobsen F, Boone T, Elliott S (1993) Role of glycosylation on the secretion and biological activity of erythropoietin. *Biochemistry* 31: 9871–9876
24. Egrie JC, Browne JK (2001) Development and characterization of novel erythropoiesis stimulating protein (NESP). *Br J Cancer* 84(Suppl 1):3–10
25. Wrighton NC, Farrell FX, Chang R, Kashyap AK, Barbone FP, Mulcahy LS, Johnson DL, Barrett RW, Jolliffe LK, Dower WJ (1996) Small peptides as potent mimetics of the protein hormone erythropoietin. *Science* 273: 458–464
26. Livnah O, Stura EA, Johnson DL, Middleton SA, Mulcahy LS, Wrighton NC, Dower WJ, Jolliffe LK, Wilson IA (1996) Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 Å. *Science* 273:464–471
27. Johnson DL, Farrell FX, Barbone FP, McMahon FJ, Tullai J, Kroon D, Freedy J, Zivin RA, Mulcahy LS, Jolliffe LK (1997) Amino-terminal dimerization of an erythropoietin mimetic peptide results in increased erythropoietin activity. *Chem Biol* 4:939–950
28. Middleton SA, Johnson DL, Jin R, McMahon FJ, Collins A, Tullai J, Gruninger RH, Jolliffe LK, Mulcahy LS (1996) Identification of a critical ligand binding determinant of the human erythropoietin receptor. Evidence for common ligand binding motifs in the cytokine receptor family. *J Biol Chem* 271:14045–14054
29. Qureshi SA, Kim RM, Konteatis Z, Biazzo DE, Motamedi H, Rodrigues R, Boice JA, Calaycay JR, Bednarek MA, Griffin P, Gao YD, Chapman K, Mark DF (1999) Mimicry of erythropoietin by a nonpeptide molecule. *Proc Natl Acad Sci U S A* 96:12156–12161
30. Goldberg J, Jin Q, Ambroise Y, Satoh S, Desharnais J, Capps K, Boger DL (2002) Erythropoietin mimetics derived from solution phase combinatorial libraries. *J Am Chem Soc* 124:544–555
31. Middleton SA, Barbone FP, Johnson DL, Thurmond RL, You Y, McMahon FJ, Jin R, Livnah O, Tullai J, Farrell FX, Goldsmith MA, Wilson IA, Jolliffe LK (1999) Shared and unique determinants of the erythropoietin (EPO) receptor are important for binding EPO and EPO mimetic peptide. *J Biol Chem* 274: 14163–14169
32. Syed RS, Reid SW, Li C, Cheetham JC, Aoki KH, Liu B, Zhan H, Osslund TD, Chirino AJ, Zhang J, Finer-Moore J, Elliott S, Sitney K, Katz BA, Matthews DJ, Wendoloski JJ, Egrie J, Stroud RM (1998) Efficiency of signalling through cytokine receptors depends critically on receptor orientation. *Nature* 395:511–516

33. Bulut GB, Sulahian R, Ma Y, Chi NW, Huang LJ (2011) Ubiquitination regulates the internalization, endolysosomal sorting, and signaling of the erythropoietin receptor. *J Biol Chem* 286: 6449–6457
34. Kamishimoto J, Tago K, Kasahara T, Funakoshi-Tago M (2011) Akt activation through the phosphorylation of erythropoietin receptor at tyrosine 479 is required for myeloproliferative disorder-associated JAK2 V617F mutant-induced cellular transformation. *Cell Signal* 23: 849–856
35. Livnah O, Johnson DL, Stura EA, Farrell FX, Barbone FP, You Y, Liu KD, Goldsmith MA, He W, Krause CD, Pestka S, Jolliffe LK, Wilson IA (1998) An antagonist peptide-EPO receptor complex suggests that receptor dimerization is not sufficient for activation. *Nat Struct Biol* 5:993–1004
36. Adams DJ (1975) Grand canonical ensemble Monte Carlo for a Lennard-Jones fluid. *Mol Phys* 29:307–311
37. Guarnieri F (2004) Computational protein probing to identify binding sites. US Patent no. 6,735,530. Issued 11 May 2004
38. Guarnieri F, Mezei M (1996) Simulated annealing of chemical potential: a general procedure for locating bound waters. application to the study of the differential hydration propensities of the major and minor grooves of DNA. *J Am Chem Soc* 118:8493–8494
39. Kulp JL 3rd, Kulp JL Jr, Pompliano DL, Guarnieri F (2011) Diverse fragment clustering and water exclusion identify protein hot spots. *J Am Chem Soc* 133:10740–10743
40. Bradbury MW, Stump D, Guarnieri F, Berk PD (2011) Molecular modeling and functional confirmation of a predicted fatty acid binding site of mitochondrial aspartate aminotransferase. *J Mol Biol* 412:412–422
41. Mezei M (1987) Grand-canonical ensemble Monte Carlo study of dense liquid Lennard-Jones, soft spheres and water. *Mol Phys* 61: 565–582
42. Kulp JL 3rd, Blumenthal SN, Wang Q, Bryan RL, Guarnieri F (2012) A fragment-based approach to the SAMPL3 Challenge. *J Comput Aided Mol Des* 26:583–594
43. Imai T, Ohura K (2010) The role of intestinal carboxylesterase in the oral absorption of pro-drugs. *Curr Drug Metab* 11:793–805
44. Zawilska JB, Wojcieszak J, Olejniczak AB (2013) Prodrugs: a challenge for the drug development. *Pharmacol Rep* 65:1–14
45. Bursi R, Grootenhuis A, van der Louw J, Verhagen J, de Gooyer M, Jacobs P, Leysen D (2003) Structure-activity relationship study of human liver microsomes-catalyzed hydrolysis rate of ester prodrugs of MENT by comparative molecular field analysis (CoMFA). *Steroids* 68:213–220
46. Krosl J, Damen JE, Krystal G, Humphries RK (1996) Interleukin-3 (IL-3) inhibits erythropoietin-induced differentiation in Ba/F3 cells via the IL-3 receptor alpha subunit. *J Biol Chem* 271:27432–27437
47. Bugelski PJ, Makropoulos D, Sprinka-Doms T, Eirikis E, Volk A, Jiao Q, Huang C (2011) Differential effects of long-lived erythropoietin receptor agonists in rats. *Pharm Anal Acta* 2:2153–2435
48. Hara KY, Mori H (2006) An efficient method for quantitative determination of cellular ATP synthetic activity. *J Biomol Screen* 11:310–317

# Chapter 15

## Designing an Orally Available Nontoxic p38 Inhibitor with a Fragment-Based Strategy

Frank Guarnieri

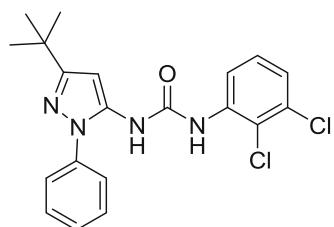
### Abstract

The MAPK p38 became a focal point of inflammatory research when it was recognized that it played a key role in the production of the pro-inflammatory molecules TNF-alpha, IL-beta, and cyclooxygenase-2 (COX-2). The pharmaceutical industry devoted enormous efforts to creating p38 inhibitors, because blocking p38 had the potential of downregulating a group of pro-inflammatory mediators, and thus, one drug could have a cocktail effect. The market potential seemed to be clearly established (Bonafede et al., Clinicoecon Outcomes Res 6:381–388, 2014) with a multiplicity of TNF-alpha antibodies and a soluble receptor (Mewar and Wilson, Br J Pharmacol 162:785–791, 2011) already on the market, although the relationship between TNF-alpha production and p38 activation is a complicated two-way (Sabio and Davis, Semin Immunol 26:237–245, 2014) signal transduction process. With the discovery that activated p38 stabilizes (Mancini and Di Battista, Inflamm Res 60:1083–1092, 2011) COX-2 mRNA and upregulates expression of IL-beta (Bachstetter and Van Eldik, Aging Dis 1:199–211, 2010) probably in a similar manner, inhibiting p38 appeared to be a way of blocking TNF-alpha, COX-2, and IL-beta simultaneously. At Locus Pharmaceuticals we jumped on this opportunity, because we believed that our fragment-based drug discovery approach was ideally suited for making a potent small molecule p38 inhibitor that did not bind in the ATP site, but also had the solubility, lack of planarity, and low molecular weight required of a clinical candidate. Just to be clear, in our experience highly planar compounds often result in poor pharmacokinetics, because they tend to bind strongly to plasma proteins. At Locus we typically repeated assays by adding increasing amounts of plasma to check for potency degradation in the presence of blood. We found this to be a useful early indicator of pharmacokinetics and in vivo behavior. It became clear from our work and the work of others that binding to the ATP site resulted in nonspecific isoform toxicities, but binding in the adjacent allosteric DFG-site resulted in molecules that were too planar and too hydrophobic. Applying the computational method of Simulated Annealing of Chemical Potential (SACP) to this problem, we at Locus were able to come up with surprising fragment substitution patterns that led to potent non-ATP p38 inhibitors with the solubility and lack of planarity that resulted in potent in vivo efficacy in rodents with 33 % oral bioavailability. By using the simulations, we made only a small number of molecules and created a high quality clinical candidate. We also did extensive co-crystallography work, which demonstrated that the compounds bound in the mode predicted by the simulations. Unfortunately, all p38 programs ultimately shut down, because compelling evidence emerged that inhibiting p38 had no long-term clinical (Genovese, Arthritis Rheum 60:317–320, 2009) benefit. Devoting a large amount of limited resources to a target that ultimately turns out to be a mistake because it was not properly validated is a fatal error for a small company, and this is one of the reasons that Locus ultimately failed.

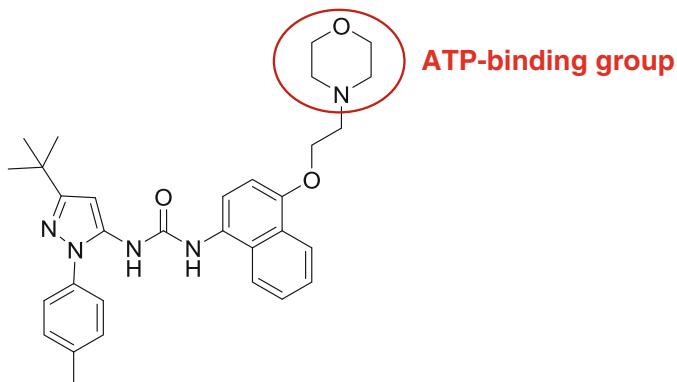
**Key words** p38, Fragment-based drug design, ATP site, DFG site, Monte Carlo, Plasma assays

## 1 Introduction

The discovery in the early 1990s that activated p38 MAP kinase induces expression of a range of pro-inflammatory cytokines made it an attractive target for pharmaceutical [1–8] drug discovery. Combinatorial chemistry and high-throughput screening at Bayer [9] led to the discovery of a class of p38 inhibitors exemplified by the molecule shown in Fig. 1. While this compound has good potency in both p38 inhibition and cellular assays, its planar hydrophobic nature is generally associated with a poor in vivo profile. Nevertheless, because it does not bind at the ATP site, it is an example of an allosteric inhibitor that binds in the pocket vacated by the kinase activation loop—specifically the highly conserved ASP-PHE-GLY motif commonly referred to as the DFG binding site. Just to be clear, the DFG sequence of the activation loop of p38 MAP kinase behaves like an “intramolecular ligand” and this class of compounds binds to and displaces the DFG triplet from the p38 allosteric site. The Boehringer Ingelheim group subsequently showed that adding a morpholino-ATP binding moiety (Fig. 2) to



**Fig. 1** Bayer p38 inhibitor that binds in the allosteric DFG site. They report that this compound has an IC<sub>50</sub> of 30 nm and inhibits p38 in SW 1353 cells at 70 nm



**Fig. 2** This is the Boehringer Ingelheim p38 inhibitor, BIRB-796, taken into human clinical trials. Note that the DFG-binding motif is essentially identical to the Bayer compound shown in Fig. 1. The di-Chloro-Phenyl group in the Bayer compound is a hydrophobic isostere of the naphthyl group in the BI compound. The new feature of the BI molecule is the morpholino group that binds in the ATP site. This makes BIRB-796 a hybrid DFG-ATP binding molecule

this class of compounds results in a subnanomolar p38 inhibitor with slow on kinetics [10] due to the displacement of the activation loop from the allosteric site. Importantly, the morpholino group adds significant solubility to the DFG-binding molecule. The BI group also reported that BIRB-796 had low affinity for 11 other kinases [11], and thus, it was deemed sufficiently selective to take into clinical trials.

---

## 2 The p38 Drug Discovery Program at Locus Pharmaceuticals

The aim of this chapter is to describe the strategy and execution of the computational fragment-based drug discovery method used to create an isoform specific, orally available, small molecule p38 inhibitor.

### 2.1 ATP or Non-ATP Binding

After intensely studying the results from the Bayer and BI groups, there was a disagreement at Locus about whether it was good or bad to be binding in the ATP site. The pro ATP-binding site camp pointed out that the Bayer molecule, which binds in the DFG site and not the ATP site is too hydrophobic and thus could never have the *in vivo* profile required of a real clinical candidate. This camp made the case that the hybrid DFG-and-ATP binding BIRB-796 not only had the needed solubility but was also highly selective since it did not appreciably bind to 11 other kinases. They were also impressed with the fact that it inhibited p38 with ~100 pm affinity. I was in the other camp and insisted that binding in the ATP-site would at some point result in off-target interactions resulting in unacceptable toxicities. My counter to the BI claim of kinase selectivity was that they never tested for p38 isoform selectivity (to our knowledge)—a viable drug-like candidate must only inhibit p38 in inflammatory cells and not inhibit p38 expressed in other tissue types. While it was not absolutely clear that this must be true at the time that this program was active, it seemed obvious and compelling to a group of us at Locus and publications years [12, 13] later validated that this was indeed the case. Specifically, the p38-gamma isoform is essential for skeletal muscle metabolic adaptation. Thus, the chronic inhibition of the alpha and beta isoforms required to treat inflammation, must also be inactive against the gamma isoform to avoid deleterious muscular side effects.

### 2.2 Let Us Test BIRB-796 to See Who Is Right

We decided that the only way to settle the dispute between the group advocating for creating a compound that binds in the ATP site to maintain solubility and the group that wanted to make a purely allosteric DFG binding compound to achieve selectivity, was to synthesize BIRB-796 and test it. By testing it we mean not just against 11 random kinases as BI did, but against a panel of 100 kinases and most importantly against all four p38 isoforms.

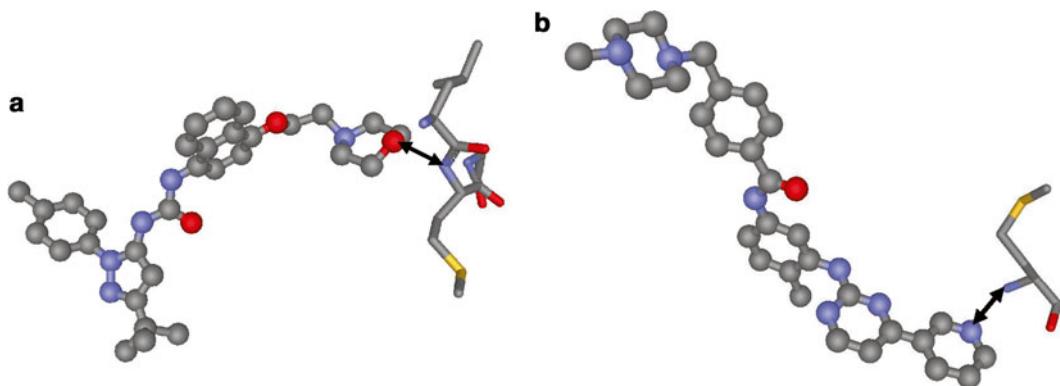
P38 Isoform	% Inhibition @ 2 $\mu$ m
P38 alpha	98
P38 beta	94
P38 delta	95
P38 gamma	77
Tie2	99
TrkB	96
JNK2 alpha2	93
C-RAF	84

**Fig. 3** As part of the Locus p38 program we synthesized and tested BIRB-796 against four different p38 isoforms and a range of kinases not tested by Boehringer Ingelheim. BIRB-796 is very potent for all p38 isoforms and a range of other kinases. One camp at Locus postulated that the ATP binding moiety of the compound was responsible for this lack of selectivity

A small but crucial subset of the data is shown in Fig. 3. We ran single point assays at a compound concentration of 2  $\mu$ m. It is important to note that this is not a dose-response curve, so this does not give an IC50. A typical procedure is to define any single point inhibition of >50 % as significant and perform a complete dose-response experiment only on these high affinity molecules. If the single point inhibition is greater than 70 % this is very significant and all of these situations are carefully analyzed with full dose-response experiments and cell assays. As the table in Fig. 3 clearly shows, BIRB-796 significantly inhibits all p38 isoforms. A small set of the other kinases that are substantially blocked is also shown. Since p38 gamma is crucial for skeletal muscle metabolic regulation, BIRB-796 cannot be given for a chronic indication such as inflammation, because continuously downregulating the p38 isoform expressed in muscle tissue is obviously unacceptable.

### 2.3 Is One Hydrogen Bond Responsible for the Non-selectivity?

The Pro-ATP-binding camp made the important point that BIRB-796 makes a key hydrogen bond in the ATP site at the so-called hinge region and that this is highly analogous to how many other small molecules bind to other kinases. The prototypical example is the binding of Gleevec to the BCR-Abl kinase. Figure 4 shows the crystal structures of BIRB-796 bound to p38 and Gleevec bound to ABL2 with this key hydrogen bond highlighted with a double-headed arrow. This leads to the compelling hypothesis that this specific interaction alone is responsible for the lack of selectivity. Just to be clear, because the hydrogen bond formed between the compounds bound in the ATP site with the amide backbone proton of a methionine is a repeating motif in kinases, the lack of selectivity may be due to the formation of this hydrogen bond alone and not to the binding in the ATP site. The Locus group



**Fig. 4** (a) The molecule in ball-and-stick is BIRB-796. The amino acid in sticks is methionine 109 in the ATP binding site from the p38 co-crystal structure 1KV2.PDB. The *black two-sided arrow* shows the very tight hydrogen bond—the distance between the oxygen from the morpholino group and the backbone nitrogen is only 2.9 Å. Note that the proton attached to this nitrogen that makes the hydrogen bond is not shown. (b) The molecule in ball-and-stick is Gleevec. The amino acid in sticks is methionine 364 in the ATP binding site from the ABL2 co-crystal structure 3GVU.PDB. The *black two-sided arrow* shows the very tight hydrogen bond—the distance between the nitrogen from Gleevec and the backbone nitrogen is only 2.9 Å. Note that the proton attached to this nitrogen that makes the hydrogen bond is not shown

that favored this hypothesis proposed replacing the morpholino group with a fragment that had high affinity for the ATP site, *but did not make this hydrogen bond*. If such a compound could be made, it would be a novel patentable derivative of BIRB-796 that retained its high potency, but also achieved the necessary alpha-beta isoform selectivity for p38 required to avoid the toxicities that will happen if the compound also bound to the delta and gamma isoforms. Even though I was in the camp that favored avoiding all ATP binding—I wanted to make purely DFG-binding molecules—I agreed with doing this experiment, because if I were wrong and simply breaking this hydrogen bond would achieve isoform selectivity, then we would have created a clinical candidate superior to BIRB-796.

### 3 Simulated Annealing of Chemical Potential (SACP)

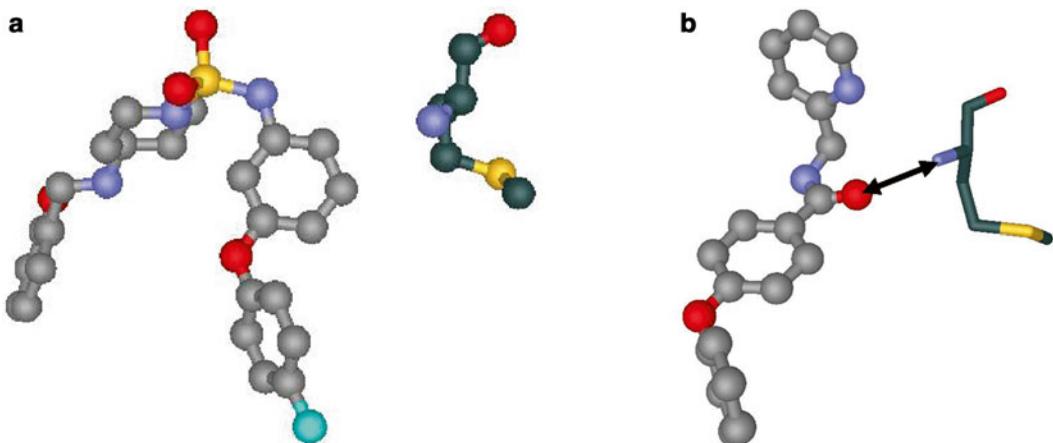
In order to discover fragments with high affinity for the ATP-binding site of p38 that do not make a hydrogen bond with the methionine in this site, SACP—the technology used to found Locus—was run on the BI crystal structure PDB.1KV2. A good simple description of the SACP method with illustrations is given in the supplementary material of a paper showing how the method located a long sought after lipid binding site [14] on the mitochondrial aspartate amino transferase. Also it is described in detail in my patent [15] on the technique, its uses [16, 17] at Locus, and

in some recent publications [18, 19] from Bioleap. There is also an abbreviated description of SACP in the chapter on making a small molecule EPO mimetic in this same volume. The first illustration of the method demonstrated the differential hydration propensities [20] between the major and minor grooves of DNA. The SACP simulations predict that diphenylether and some its derivatives will bind in the ATP site of p38 and that this fragment does not make a hydrogen bond at the hinge region. Locus synthesized several of these compounds and they were quite active in p38 inhibition assays. There is an especially interesting comparative aspect between two different classes of these compounds—the diphenylether is quite removed from the MET 109 in both cases, but in one class an amide backbone from the inhibitor makes a hydrogen bond with MET 109 while the other does not. Thus, from the SACP simulations and the Locus experiments, two closely related diphenylether molecules block p38 with comparable affinity but have different binding modes in the ATP site. *This gave Locus a definitive answer—we needed to be out of the ATP site, because whether or not the compounds H-bond to MET 109 made no difference—all of these ATP-binding compounds inhibited all four p38 isoforms with comparable affinity.* It was quite gratifying that when co-crystal structures of these compounds were made with p38, the compounds were found to bind exactly as predicted from the simulations. The co-crystals did beautifully confirm the computational predictions, so Locus published [21] the results (some key aspects of the co-crystals are shown in Fig. 5).

---

#### 4 The Challenge of Obtaining p38 Inhibition in Blood

The fragment-based experiments discussed above united the team at Locus on the fact that we had to make a purely DFG binding compound and that we needed to stay out of the ATP site if we were to create an isoform selective nontoxic clinical candidate. We restarted the program beginning with the Bayer molecule shown in Fig. 1. Our working hypothesis was that this molecule needed to be solubilized with a heterocyclic aliphatic functional group. Clearly BI was on the right track that this molecule required a solubilizing moiety, but it also required something that would break the planarity of the compound. We believed that this was necessary for maintaining the potency of the compound when plasma was added to the assays. Just to be clear, in our experience highly planar compounds often have a strong propensity for tightly binding to blood proteins. To illustrate this point we performed a triplet set of assays on the BIRB-796, which is shown in Fig. 2. Our assays confirmed that BIRB-796 is a very potent p38 inhibitor. When we used the full length MKK in the phosphorylation assays, the IC<sub>50</sub> was around 2 nm. BI used a peptide fragment of MKK, which likely accounts for the small discrepancy—BI reported a subnanomolar

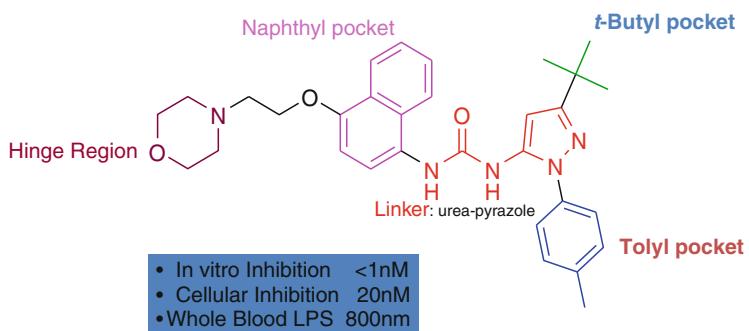


**Fig. 5** (a) The molecule in ball-and-stick is a new diphenylether class of p38 inhibitors discovered using the computational fragment-based drug discovery technique “Simulated Annealing of Chemical Potential” or SACP. Methionine 109 is shown in ball-and-stick with the carbons colored in slate gray. This is from co-crystal structure 1ZZ2.PDB of p38 with the new class of compounds created at Locus experimentally confirming the computational prediction of diphenylether binding in the ATP site, but NOT making a hydrogen bond with the backbone of MET 109. (b) The molecule in ball-and-stick is a second new diphenylether class of p38 inhibitors discovered using SACP. Methionine 109 is shown in sticks with the carbons colored in slate gray. This is from co-crystal structure 1ZYJ.PDB of p38 with the new class of compounds created at Locus. Interestingly, an amide carbonyl from the molecule makes a tight hydrogen bond with the backbone amide of MET 109—this is highlighted with the *two-headed arrow*. The oxygen-nitrogen distance is only 2.9 Å. Note the amide proton making the H-bond is not shown

IC<sub>50</sub>. We then used a THP cell line. THP cells have been used for ~30 years. They were derived from a patient [22] that had acute monocytic leukemia. The singular benefit of using this cell line is that it only requires 0.5 % serum in the media to maintain its normal activity. The reason that this is so important is because the compounds are tested for their ability to cross the cell membrane without the confounding factor of whether or not they are sequestered by the blood. The EC<sub>50</sub> of BIRB-796 in the THP assays was 20 nm. The third assay used a PBMC cell line that required 5 % serum in order to get a preliminary indication of how the compound behaves in the presence of blood. The three step assay protocol is, (1) obtain the IC<sub>50</sub> for directly inhibiting p38, (2) because p38 resides inside the cell, obtain the EC<sub>50</sub> of inhibition using the THP cancer cell line that requires virtually no serum as a test for how well the compound crosses the cell membrane, (3) repeat the cell experiments using PBMC cells, because they require 5 % serum and thus comparing the difference between the 2 cell assays give a first indication of how well the inhibitor behaves in the presence of blood. A final in vitro (sometimes called ex vivo) assay is the whole blood LPS test. In this assay whole blood is taken from a donor. When bacterial lipopolysaccharide, a component of the gram negative outer membrane, is added to the blood it elicits a violent inflammatory reaction resulting in the a

dramatic upregulation of TNF. The p38 inhibitor is then added and the decrease in TNF production is monitored. The whole blood LPS assay is reserved for the best compounds, because it can be difficult to perform. The variations of TNF in whole blood exposed to LPS can be very large so a great deal of standardization may be required. This demonstrates one of the practical difficulties encountered in drug discovery. In theory having an *in vitro* assay that is a reliable indicator of *in vivo* behavior is highly desirable. Often the high degree of difficulty in performing such an assay leads to the conclusion that it is just easier to go straight into the animal and get the *in vivo* data directly. The results of the triplet assay on the BI compound are summarized in Fig. 6.

While it is imperative to remove the morpholino group from BIRB-796 in order to achieve isoform selectivity, there is the additional extremely positive benefit of reducing the molecular weight by 120 Da. Oprea [23] emphasizes that quality lead compounds *in general* are low molecular weight molecules ( $MW < 425$ ) with good solubility and modest  $\log P$  values (<3). Just to be clear, Oprea does not say that there are not some exceptions to this characterization, but that overall there is a better chance of developing a high quality clinical candidate if these constraints are approximately adhered to. Thus, after intensive study of the p38 literature and our own work, we concluded that for a compound to be a viable clinical candidate it had to bind exclusively in the DFG site with no ATP-binding for isoform selectivity (avoid toxicity), and have significant solubility coming from a new and different three-dimensional heterocyclic functional group (maintain potency in plasma). Our goal was to create a low molecular weight (~450 Da), nonplanar, soluble, DFG binding compound that inhibits p38



**Fig. 6** A summary of the triplet assay procedure used at Locus in the p38 program. Our working hypothesis was that it would be much better to create an inhibitor with substantially lower MW that had a more consistent efficacy from protein inhibition to cell inhibition to cell inhibition in the presence of blood. In other words, pm inhibition in a protein assay with a compound that loses four orders of magnitude efficacy when progressively going to more biologically relevant assays is probably an indication of something that will have a poor *in vivo* profile

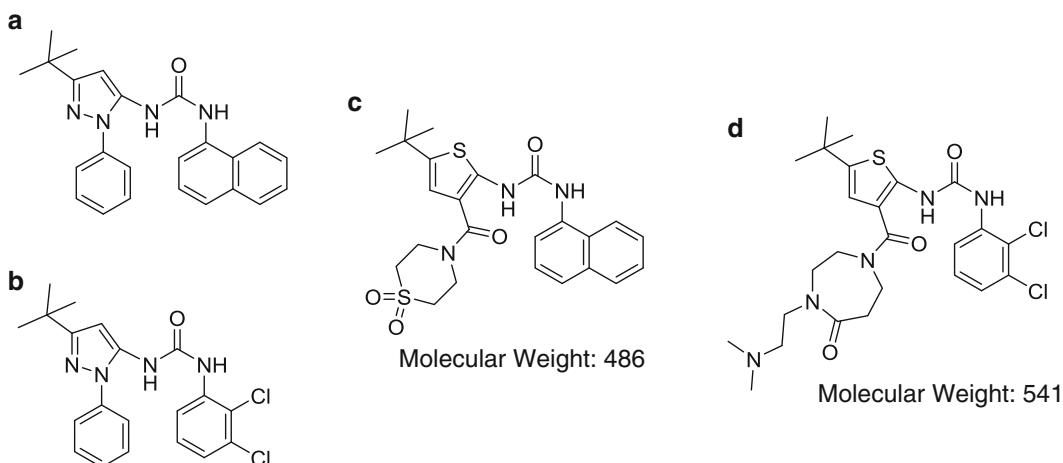
with submicromolar efficacy in a cell assay in the presence of significant plasma. Our hypothesis was that such a molecule would be orally bioavailable and have significant in vivo potency. Just to be clear, the singular focus on making compounds that bind to p38 that are nanomolar to subnanomolar without worrying about anything else, seems to result in molecules that either behave poorly in blood or bind to the ATP site and thus lack the isoform specificity required to be nontoxic. We felt that striving for submicromolar in a cell assay in the presence of blood was a better predictor of a good in vivo profile.

#### **4.1 Can the Highly Conserved Phenyl Group Be Replaced?**

One of the functional groups in the Bayer compound must be replaced if the resultant molecule is to have a molecular weight of 450 or less—obviously just adding an additional functional group as BI did would cause an unacceptable increase in the molecular weight. Bayer has a phenyl group attached to the pyrazine and BI slightly derivatizes it to the toluene. This phenyl moiety mimics the conserved phenylalanine of the DFG triplet from the activation loop that binds in the p38 allosteric site. Because this phenylalanine is so critical in the functioning of the protein, all molecules published at the time we were doing the work contained this phenyl group—it was considered essential. The SACP simulations made the startling predictions that certain complicated aliphatic heterocycles had approximately the same free energy of binding as the phenyl group and the toluene fragments. I say that this was startling, because at the time it was generally accepted that this was a hydrophobic pocket designed by nature to bind the benzene ring of the conserved DFG triplet. The two fragments that we decided to incorporate into the Bayer scaffold are shown in Fig. 7. It was very gratifying that co-crystal structures of these compounds showed that they bound exactly as predicted by the simulations. The co-crystal structures and how they confirm and validate the computational predictions have been published by Locus [24] and many of the structures have been deposited in the protein data bank.

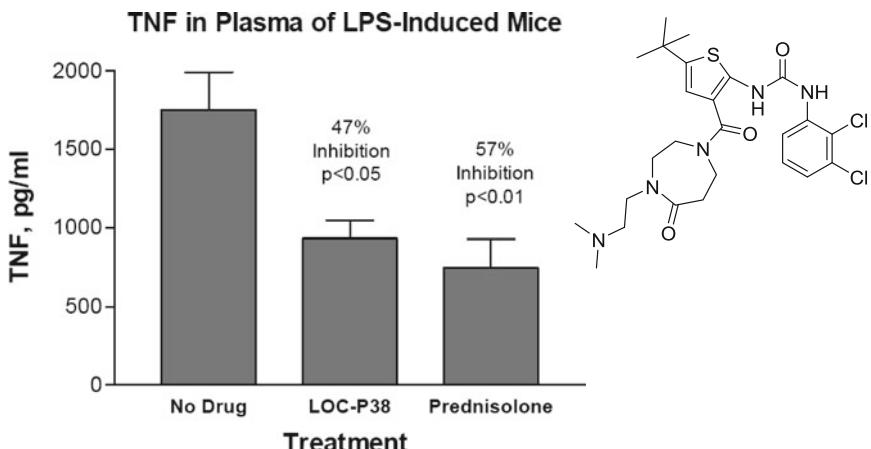
#### **4.2 The Importance of Activity in Blood as an Indicator of In Vivo Activity**

SACP simulations predict the relative rank order binding affinity of different fragments for a given pocket in the protein, but have nothing to say about how the compounds will behave in the presence of blood. Our studies of albumin binding profiles indicated that planar structures have a higher probability of being sequestered by blood proteins. In fact, in our experiments, BIRB-796 is 99 % plasma protein bound. Once we confirmed the SACP simulations predicting that dioxothiomorpholine and diazepanone could be substituted for the phenyl group by inhibition assays and co-crystallography, we made a set of derivatives of the two fragments and assayed their potency in the presence of high amounts of plasma. One compound stood out as maintaining sub-



**Fig. 7** Compounds A and B are the core structures discovered by Bayer and BI. Note that the dichlorophenyl moiety in B is considered to be a hydrophobic isostere of the naphthyl group in A. The SACP simulations predict that thiophene can be substituted for pyrazine. SACP also predicts that the dioxothiomorpholine shown in compound C and the diazepanone shown in compound D are equipotent replacements of the phenyl group. Locus medicinal chemists were adamant that no one would have ever thought to make these substitutions to create compounds C and D, because the phenyl binding pocket was considered to be the conserved aspect of the DFG activation loop intramolecular interaction and thus must be kept

micromolar inhibition in a cell assay even when high amounts of human plasma was added to the media—this is compound D shown in Fig. 7. Accordingly we tested this molecule *in vivo* showing that it has 33 % oral bioavailability in a rat pharmacokinetic model and that it has significant efficacy in blocking p38 induced TNF production in a mouse inflammatory [25] model (shown in Fig. 8). These results validate our strategy for creating a p38 clinical candidate, which can be summarized as, (1) no ATP binding to avoid isoform nonspecific toxicity, (2) achieving solubility by replacing a hydrophobic moiety generally thought to be essential with a soluble aliphatic heterocycle predicted from the SACP simulations, (3) breaking the planarity of the molecule to retain a good deal of the potency in the presence of blood, (4) integrating the simulation results with a triplet assay of binding to p38, p38 inhibition in THP cells, and p38 inhibition in PBMC cells. While our compounds did not bind with subnanomolar potency to p38—they were generally about 20 nm—they were designed to be sub-micromolar in cell assays in the presence of blood. We felt that maintaining good potency when going to successively more biologically relevant assays was a better indicator of what would happen *in vivo* compared to having extreme potency in a binding assay with orders of magnitude of degradation in potency when going to cell assays and then plasma assay. Our *in vivo* results of 33 % oral bioavailability and significant p38 blockage in rodent models demonstrates that this approach at least for a p38 inhibitor is a good way to achieve the *in vivo* results required of a clinical candidate.



**Fig. 8** Compound D from Fig. 7 was taken into animal studies, because it inhibited p38 with an IC<sub>50</sub> of 20 nm and it inhibited p38 in a PBMC cell line that requires 5 % serum with 400 nm EC<sub>50</sub>, and most importantly was still 860 nm in the PBMC assay when large quantities of human plasma was added. This compound has 33 % oral bioavailability in a rat PK model and is highly selective for the p38alpha isoform and shows significant p38 inhibition in vivo

## 5 Conclusions

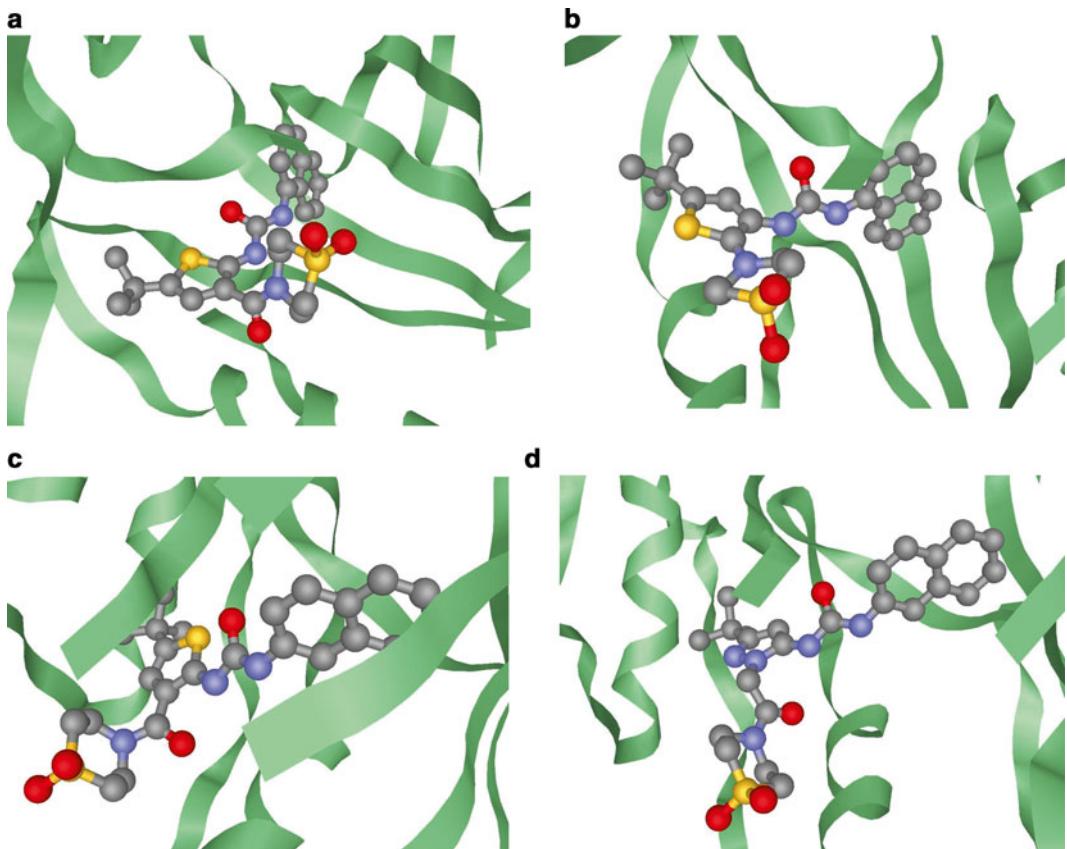
Unfortunately, the p38 program at Locus was ultimately a failure because Cohen and coworkers [26] showed that an isoform selective p38 inhibitor was less effective than methotrexate in treating active rheumatoid arthritis. Genovese [6] nicely summarizes these results in an editorial by stating, “The p38 MAP kinase inhibitors represent one of the most heralded classes of therapies for the treatment of inflammation in the past decade,” and “the fat lady has sung for the beleaguered development of p38 inhibitors for the treatment of RA.” The business decision at Locus to put so many resources into this program along with other questionable business decisions resulted in the company going bankrupt after about 10 years in existence. *The main lesson for a small company is – make sure that your target is clinically validated.* As an example, blocking PCSK9—the liver protein that downregulates the LDL receptor has been clinically validated as an anticholesterol target. Amgen has excellent data [27–30] in Phase III clinical trials with an anti-PCSK9 antibody, and thus, this is a validated high value target for small molecule development. A small company simply cannot afford to make the mistake that we made at Locus—getting caught up in the current hype surrounding a new and exciting target without first seeing validation in a clinical setting. The target must be proven in humans and you must have a strategy for why you can do better scientifically and *more importantly why you can do better economically*—ultimately this is business and getting overly caught up in the science was a big mistake.

---

## 6 Final Thoughts and Acknowledgements

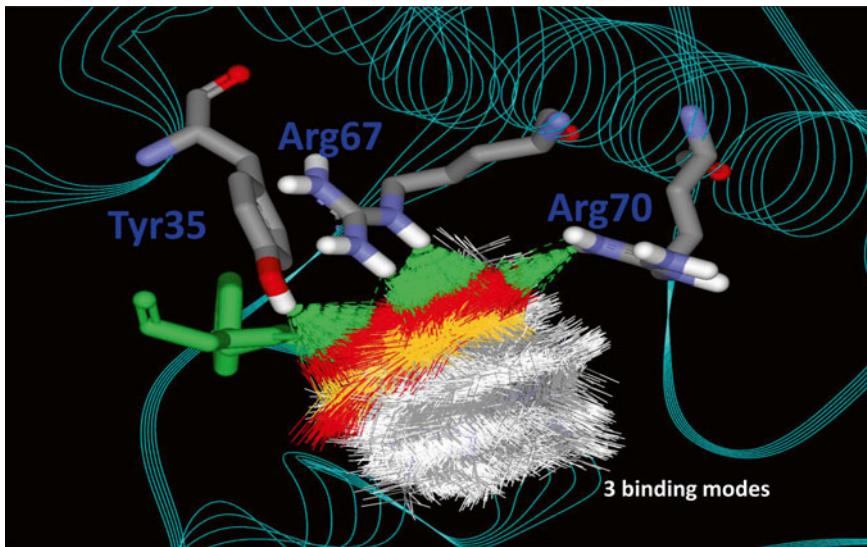
At the time that the work was going on I was caught up in proving that the SACP simulations that I invented would lead to the discovery of classes of inhibitors that no medicinal chemist would have come up with and that my ideas on non-ATP binding and tightly coupling inhibition assays with cell and blood assays would lead to a clinical candidate making only a small number of molecules. Just to be clear, several other Locus scientists felt the same way about non-ATP binding and the need for activity in human plasma, so these were not my ideas alone. Many of us felt that optimizing a compound to get nanomolar or subnanomolar potency in a binding assay was probably in fact counterproductive, since it was likely to lead to poor behavior in vivo and that the best and easiest surrogate for a good in vivo profile was maintaining potency in the presence of human blood. Of course if the target is intracellular like p38, it is vitally important to not lose too much potency when going from a protein inhibition assay to a cell assay. *We felt that a good rule of thumb would be to lose not more than one order of magnitude in potency when going from a protein inhibition assay to a cell assay and not more than an additional one order of magnitude in potency when blood was added to the cell assay.* Basically, we thought that losing many orders of magnitude of potency when progressing to more biologically relevant assays was a bad omen for in vivo efficacy. The SACP predictions that nonplanar heterocyclic functionalities could replace the conserved benzene ring of the DFG triplet of the activation loop was a completely startling result and I have to thank the Locus chemists for being open-minded enough to synthesize the molecules and the Locus biologists for doing the assays and obtaining co-crystals. The Locus chemists went even further—the SACP simulations predicted that the dioxothiomorpholine could bind in several different modes and they actually synthesized a few variants to test these computational predictions. Co-crystals of these different variants were obtained and deposited in the protein data bank validating the computational predictions.

Figure 9a–d show the co-crystal structures of the Locus p38 inhibitors with the dioxothiomorpholine fragment in place of the phenyl group that was generally accepted as irreplaceable. The SACP simulations predicted that the dioxothiomorpholine fragment would bind in three different modes (Fig. 10) if a one carbon linker was used. Referring back to Figs. 1 and 2, both Bayer and BI directly connected the phenyl ring to the pyrazine scaffold. The SACP simulations predicted that directly linking the dioxothiomorpholine fragment to the pyrazine or thiophene scaffold would also be fine, but would be confined to one binding mode. The four co-crystals shown in Fig. 9 experimentally confirm these computational predictions.

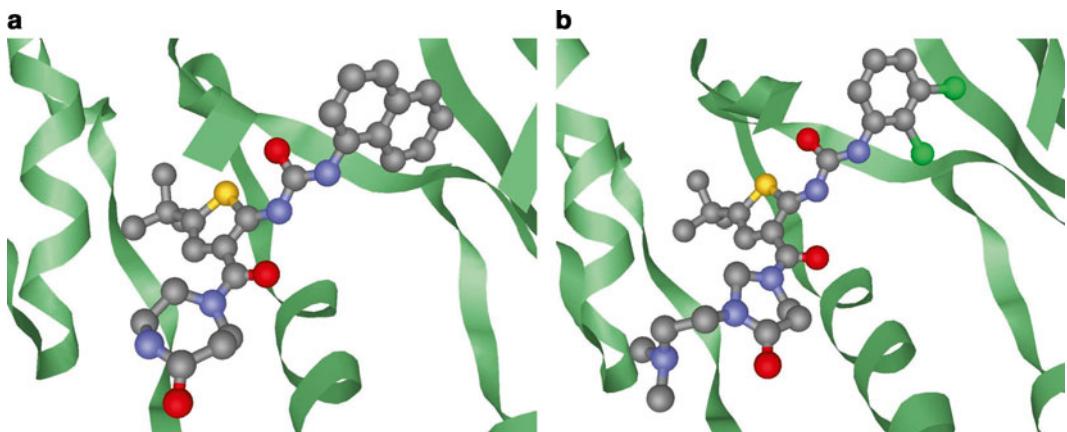


**Fig. 9** (a) A binding mode of the dioxothiomorpholino p38 inhibitor with a one carbon link to the thiophene scaffold taken from 3P5K.PDB experimentally confirming the computational prediction from SACP simulations. (b) A different binding mode of the dioxothiomorpholino p38 inhibitor taken from 3P7A.PDB. The SACP simulations predicted that an alternate binding mode would occur if the dioxothiomorpholine fragment was directly linked to the thiophene scaffold and co-crystallography confirms this computational prediction. (c) The SACP simulations predicted that the one carbon linker to the thiophene would result in multiple binding modes of the dioxothiomorpholine fragment. This co-crystal is taken from 3P78.PDB, which experimentally confirms the prediction of the additional binding modes. (d) The SACP simulations actually predicted three different binding modes for the dioxothiomorpholine fragment with a one carbon linker to the thiophene. This co-crystal is taken from 3P79.PDB, which experimentally confirms this third binding mode

Figure 11a, b shows the co-crystal structures of the diazepanone fragment predicted from SACP to be a replacement for the conserved phenylalanine. We considered the molecule in Fig. 11b to be our best candidate for clinical trials, because it inhibits p38 in a direct binding assay at 20 nm, inhibits p38 in a PBMC cell line in 5 % serum at 400 nm, and inhibits p38 in the PBMC cell line with added human plasma at 800 nm. Even though this compound is not a 100 pm inhibitor in a direct binding assay, the fact that it lost only one order of magnitude in potency when going from a protein inhibition assay to a cell assay in serum and then lost virtually no additional potency when human serum was added was taken as



**Fig. 10** SACP predicts that the dioxothiomorpholine fragment will bind with high affinity in the conserved phenylalanine binding DFG pocket. Interestingly, all three of these binding modes are predicted to occur for a 1-carbon link to the pyrazine or thiophene scaffold, and only one of these modes is predicted to occur with a direct link to the pyrazine or thiophene scaffold. Co-crystal structures shown in Fig. 9 confirm these predictions



**Fig. 11** (a) This co-crystal is taken from 3P7B.PDB, which experimentally confirms the SACP prediction that the diazepanone fragment can replace the phenyl group that replaces the F in the DFG triplet of the activation loop. (b) This co-crystal is taken from 3P7C.PDB. SACP simulations indicated that the diazepanone fragment could be derivatized at this position with alkyl amines. This, however was more of a Medicinal Chemistry effort to obtain better affinity when human plasma was added to the cell assays

an indication that the compound had a high likelihood of behaving well *in vivo*. The animal data show (Fig. 8) that this molecule has high oral bioavailability and significant *in vivo* efficacy, supporting the hypothesis that submicromolar efficacy in blood is a good indicator of *in vivo* potency.

One final comment—this molecule violates our initial desire to keep the molecular weight low—hopefully below 450 but at least under 500 Da. The dioxothiomorpholine compounds were designed with this goal in mind, but for some unknown reason this fragment resulted in compounds that lost too much efficacy in the presence of human plasma. We really wanted the compound taken into *in vivo* studies to be submicromolar in the presence of human plasma. This required relaxing the low molecular weight constraint. The *in vivo* data confirm the benefits of seeing submicromolar activity in the presence of human blood and the goal of making a compound that maintains most of its potency when progressing through a set of assays that are increasingly biologically relevant.

## References

- Bonafede M, Joseph GJ, Shah N, Princic N, Harrison DJ (2014) Cost of tumor necrosis factor blockers per patient with rheumatoid arthritis in a multistate Medicaid population. *Clinicoecon Outcomes Res* 6:381–388
- Mewar D, Wilson AG (2011) Treatment of rheumatoid arthritis with tumour necrosis factor inhibitors. *Br J Pharmacol* 162:785–791
- Sabio G, Davis RJ (2014) TNF and MAP kinase signalling pathways. *Semin Immunol* 26:237–245
- Mancini AD, Di Battista JA (2011) The cardinal role of the phospholipase A(2)/cyclooxygenase-2/prostaglandin E synthase/prostaglandin E(2) (PCPP) axis in inflammostasis. *Inflamm Res* 60:1083–1092
- Bachstetter AD, Van Eldik LJ (2010) The p38 MAP kinase family as regulators of proinflammatory cytokine production in degenerative diseases of the CNS. *Aging Dis* 1:199–211
- Genovese MC (2009) Inhibition of p38: has the fat lady sung? *Arthritis Rheum* 60:317–320
- Han J, Lee JD, Bibbs L, Ulevitch RJ (1994) A MAP kinase targeted by endotoxin and hyperosmolarity in mammalian cells. *Science* 265:808–811
- Lee JC, Laydon JT, McDonnell PC, Gallagher TF, Kumar S, Green D, McNulty D, Blumenthal MJ, Heys JR, Landvatter SW et al (1994) A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature* 372:739–746
- Dumas J, Hatoum-Mokdad H, Sibley R, Riedl B, Scott WJ, Monahan MK, Lowinger TB, Brennan C, Natero R, Turner T, Johnson JS, Schoenleber R, Bhargava A, Wilhelm SM, Housley TJ, Ranges GE, Shrikhande A (2000) 1-Phenyl-5-pyrazolyl ureas: potent and selective p38 kinase inhibitors. *Bioorg Med Chem Lett* 10:2051–2054
- Regan J, Pargellis CA, Cirillo PF, Gilmore T, Hickey ER, Peet GW, Proto A, Swinamer A, Moss N (2003) The kinetics of binding to p38MAP kinase by analogues of BIRB 796. *Bioorg Med Chem Lett* 13:3101–3104
- Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, Grob PM, Hickey ER, Moss N, Pav S, Regan J (2002) Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat Struct Biol* 9:268–272
- Long DL, Loeser RF (2010) p38gamma mitogen-activated protein kinase suppresses chondrocyte production of MMP-13 in response to catabolic stimulation. *Osteoarthritis Cartilage* 18:1203–1210
- Pogozelski AR, Geng T, Li P, Yin X, Lira VA, Zhang M, Chi JT, Yan Z (2009) p38gamma mitogen-activated protein kinase is a key regulator in skeletal muscle metabolic adaptation in mice. *PLoS One* 4:e7934
- Bradbury MW, Stump D, Guarnieri F, Berk PD (2011) Molecular modeling and functional confirmation of a predicted fatty acid binding site of mitochondrial aspartate aminotransferase. *J Mol Biol* 412:412–422

15. Guarnieri F (2004) Computational protein probing to identify binding sites. US Patent No. 6,735,530. Issued 11 May 2004
16. Clark M, Guarnieri F, Shkurko I, Wiseman J (2006) Grand canonical Monte Carlo simulation of ligand-protein binding. *J Chem Inf Model* 46:231–242
17. Moore WR Jr (2005) Maximizing discovery efficiency with a computationally driven fragment approach. *Curr Opin Drug Discov Devel* 8:355–364
18. Kulp JL III, Blumenthal SN, Wang Q, Bryan RL, Guarnieri F (2012) A fragment-based approach to the SAMPL3 challenge. *J Comput Aided Mol Des* 26:583–594
19. Kulp JL III, Kulp JL Jr, Pompliano DL, Guarnieri F (2011) Diverse fragment clustering and water exclusion identify protein hot spots. *J Am Chem Soc* 133:10740–10743
20. Guarnieri F, Mezei M (1996) Simulated annealing of chemical potential: a general procedure for locating bound waters. Application to the study of the differential hydration propensities of the major and minor grooves of DNA. *J Am Chem Soc* 118:8493–8494
21. Michelotti EL, Moffett KK, Nguyen D, Kelly MJ, Shetty R, Chai X, Northrop K, Namboodiri V, Campbell B, Flynn GA, Fujimoto T, Hollinger FP, Bukhtiyarova M, Springman EB, Karpasas M (2005) Two classes of p38alpha MAP kinase inhibitors having a common diphenylether core but exhibiting divergent binding modes. *Bioorg Med Chem Lett* 15: 5274–5279
22. Qin Z (2012) The use of THP-1 cells as a model for mimicking the function and regulation of monocytes and macrophages in the vasculature. *Atherosclerosis* 221:2–11
23. Oprea TI (2002) Current trends in lead discovery: are we looking for the appropriate properties? *Mol Divers* 5:199–208
24. Moffett K, Konteatis Z, Nguyen D, Shetty R, Ludington J, Fujimoto T, Lee KJ, Chai X, Namboodiri H, Karpasas M, Dorsey B, Guarnieri F, Bukhtiyarova M, Springman E, Michelotti E (2011) Discovery of a novel class of non-ATP site DFG-out state p38 inhibitors utilizing computationally assisted virtual fragment-based drug design (vFBDD). *Bioorg Med Chem Lett* 21:7155–7165
25. Zhou JS, Xing W, Friend DS, Austen KF, Katz HR (2007) Mast cell deficiency in Kit(W-sh) mice does not impair antibody-mediated arthritis. *J Exp Med* 204:2797–2802
26. Cohen SB, Cheng TT, Chindalore V, Damjanov N, Burgos-Vargas R, Delora P, Zimany K, Travers H, Caulfield JP (2009) Evaluation of the efficacy and safety of pamapimod, a p38 MAP kinase inhibitor, in a double-blind, methotrexate-controlled study of patients with active rheumatoid arthritis. *Arthritis Rheum* 60:335–344
27. Stroes E, Colquhoun D, Sullivan D, Civeira F, Rosenson RS, Watts GF, Bruckert E, Cho L, Dent R, Knusel B, Xue A, Scott R, Wasserman SM, Rocco M (2014) Anti-PCSK9 antibody effectively lowers cholesterol in patients with statin intolerance: the GAUSS-2 randomized, placebo-controlled phase 3 clinical trial of evolocumab. *J Am Coll Cardiol* 63:2541–2548
28. Robinson JG, Nedergaard BS, Rogers WJ, Fialkow J, Neutel JM, Ramstad D, Somaratne R, Legg JC, Nelson P, Scott R, Wasserman SM, Weiss R (2014) Effect of evolocumab or ezetimibe added to moderate- or high-intensity statin therapy on LDL-C lowering in patients with hypercholesterolemia: the LAPLACE-2 randomized clinical trial. *JAMA* 311: 1870–1882
29. Koren MJ, Lundqvist P, Bolognese M, Neutel JM, Monsalvo ML, Yang J, Kim JB, Scott R, Wasserman SM, Bays H (2014) Anti-PCSK9 Monotherapy for Hypercholesterolemia: The MENDEL-2 Randomized, Controlled Phase III Clinical Trial of Evolocumab. *J Am Coll Cardiol* 63:2531–2540
30. Blom DJ, Hala T, Bolognese M, Lillestol MJ, Toth PD, Burgess L, Ceska R, Roth E, Koren MJ, Ballantyne CM, Monsalvo ML, Tsirtonis K, Kim JB, Scott R, Wasserman SM, Stein EA (2014) A 52-week placebo-controlled trial of evolocumab in hyperlipidemia. *N Engl J Med* 370:1809–1819

# INDEX

## A

- Absorption, distribution, metabolism, and excretion (ADME).....67, 129, 131–132, 163, 178  
ADME toxicity (ADME/Tox).....67  
Agonist.....188–192, 197–200, 203, 207, 208  
AKT. *See* Kinase, protein kinase B  
ALIS.....44  
AMBER.....6–11, 17, 33, 37, 38, 67, 97  
ANCHOR.....60, 64  
Antagonist.....198–200, 207  
*Apo* structure.....146, 148  
Aranesp.....188, 207

## B

- BEI. *See* Binding efficiency index (BEI)  
Binding  
affinity ..... 4, 10, 11, 33, 37, 76, 78, 82, 102, 113, 114, 128, 138, 147, 149, 152, 163, 189, 197, 203, 219  
free energy ..... 4–6, 12, 18, 76, 77, 82, 90, 96, 127  
site ..... 5, 10, 12–20, 32, 34, 58, 61–66, 69, 78, 80, 90, 98, 120, 122, 123, 125, 129, 131, 132, 138, 143, 145–153, 161, 164, 167–169, 178, 179, 192, 193, 195–199, 212, 213, 215  
Binding efficiency index (BEI).....90, 91, 93–95  
Binding site analysis (BSA).....145–153, 201  
Bioisosteres.....61, 65, 66, 169  
Blood–brain barrier .....67, 178  
B3LYP functional.....37  
BSA. *See* Binding site analysis (BSA)

## C

- CACTVS .....16  
Cambridge structural database (CSD) .....168  
Cell lines  
Baf3 .....207  
THP .....217  
UT-7 .....207  
cFBDE. *See* Computational fragment based design evolution (cFBDE)  
cGRILL .....91–93, 95–97  
CHARMm force field.....63  
CHelpG single-point calculation .....37  
ChEMBL.....44

- CHEMGENIE.....44, 46–49, 51, 52  
CHEMGRID .....17  
Chemical diversity.....32, 58, 120, 126, 165, 196  
Chemical potential .....33, 78, 147, 193–195, 197, 198, 206, 215–217  
Chemical space .....17, 32, 33, 39, 44, 58, 78, 120, 122  
Cheminformatics .....43–52  
Chemogenomics .....44, 51, 125  
Chemotype .....32, 34, 58, 131, 140–143, 158, 177  
Computational fragment based design  
evolution (cFBDE) .....140–143

## D

- DelPhi .....6, 8, 11, 17  
De novo design .....89–98, 124, 127, 139, 141–143, 167–169, 171  
Descriptors .....14, 19, 69, 126, 132, 148  
molecular .....122  
Dihydrofolate reductase (DHFR) .....147, 148  
DISTMAP .....16  
Docking  
software .....5, 15, 138, 169  
AutoDock .....11, 60, 66–67, 101–113  
DOCK .....15–17, 60, 69  
FLEXX .....138  
fragment flexible ligand docking  
(FFLD) .....60, 69  
Glide XP .....5  
GOLD .....69, 138  
MOE .....69, 138, 169

## 3D-QSAR

- programs .....128  
CoMFA .....128  
CoMSIA .....128

## 3D-RISM

## E

- Energy minimization .....17, 62, 66, 68, 138  
Powell .....68  
Enthalpy .....5, 8, 195  
Entropy .....5, 7, 8, 33, 128, 138, 147, 163, 195  
Erythropoietin (EPO) .....185–208  
receptor (EPOR) .....188, 190–192, 196–200, 207, 208

## F

- Factor Xa ..... 5  
 FBDD. *See* Fragment based drug design (FBDD)  
 FEP. *See* Free energy perturbation (FEP)  
 Fingerprints (FPs)  
     daylight ..... 122  
     extended connectivity fingerprints (ECFPs) ..... 122  
     MD keys ..... 122  
     Unity 2D ..... 122  
 FK506 binding protein (FKBP) ..... 102–114  
 fpocket ..... 146  
 FragMap ..... 77, 78, 80–84, 123  
 Fragment  
     binding ..... 32, 33, 77, 119, 127, 149, 152, 178, 194, 196–198  
     database ..... 15, 34, 35, 90, 98, 124, 129, 168, 169  
     growing ..... 126, 168, 175, 178  
     library ..... 14–16, 43–52, 61, 65, 90, 93, 98, 119–122, 128, 168, 172, 175, 176  
     linking ..... 51, 57–59, 95, 123, 127–129  
     screening ..... 14–17, 19, 33, 39, 57, 58, 68, 126, 128, 132, 157, 165–167  
 Fragment based drug design (FBDD) ..... 13–20, 31–39, 57, 75–84, 89–91, 101, 120, 121, 127, 133, 137–144, 157, 161, 167, 170, 172, 175  
 Fragment based ligand discovery (FBLD) ..... 23, 161–164, 170, 172, 174–179  
 Fragment-based Virtual Library (FragVLib) ..... 23–29, 124  
 Fragment hopping ..... 57–69  
 Fragments of active structures (FAST) ..... 123  
 FragVLib. *See* Fragment-based Virtual Library (FragVLib)  
 Free energy  
     binding ..... 4–6, 12, 18, 75–77, 90, 96, 127  
     desolvation ..... 17, 19, 77  
 Free energy perturbation (FEP) ..... 4, 5, 138  
 FTMap ..... 64, 68, 123

## G

- 6–31G(d) basis set ..... 37, 38  
 GCMC. *See* Grand canonical Monte Carlo (GCMC)  
 Generalized AMBER force field (GAFF) ..... 97, 98  
 GFE. *See* Grid free energy (GFE)  
 GOLPE ..... 60, 63, 64, 69  
 Grand canonical Monte Carlo (GCMC) ..... 33, 139, 147, 152, 167, 193  
 GRID ..... 61, 63, 64, 68, 69, 91, 122, 123, 126  
 Grid free energy (GFE) ..... 75, 81, 84  
 Grid Inhomogeneous Solvation Theory (GIST) ..... 5  
 GRIN ..... 61  
 GROMACS ..... 9

## H

- HBA. *See* Hydrogen bond acceptor (HBA)  
 HBD. *See* Hydrogen bond donor (HBD)  
 High concentration assay (HCA) ..... 138  
 High throughput screening (HTS) ..... 32, 44, 51, 58, 76, 120, 212  
 HIV. *See* Human immunodeficiency virus (HIV)  
 Holo structure ..... 146  
 HTS. *See* High throughput screening (HTS)  
 Human immunodeficiency virus (HIV) ..... 8, 9, 148  
 Hydration free energy (HFE) ..... 4, 7  
 Hydrogen bond acceptor (HBA) ..... 66, 77, 80, 81, 91, 92, 95, 97, 98, 147, 150  
 Hydrogen bond donor (HBD) ..... 34, 66, 77, 80, 81, 89, 91–93, 97, 141, 147, 163, 176

## I

- IMPACT ..... 17  
 Implicit solvation ..... 4, 6  
 Implicit solvent model (ISM) ..... 6, 16, 95  
 InChIKey ..... 44, 47, 48, 51  
 Inhomogeneous fluid solvation theory (IFST) ..... 5, 7, 8  
 Integral equation theories (IETs) ..... 7  
 ISM. *See* Implicit solvent model (ISM)  
 Isothermal titration calorimetry (ITC) ..... 76, 120, 138, 164, 165

## J

- Jaccard coefficient ..... 122

## K

- Kinases  
     ATP pocket ..... 158, 159, 161  
     BCR-Abl ..... 214  
     BRAF ..... 179  
     CDK2 ..... 168, 169, 172, 174, 175  
     CHK1 ..... 174, 177  
     CSK ..... 132  
     DFG loop ..... 161, 169  
     FLT3 ..... 175, 177  
     GSK3 $\beta$  ..... 175, 177  
     hinge ..... 39, 160, 166, 170, 171  
     inhibitors ..... 35, 157–179, 221  
         BIRB-796 ..... 35  
         design of ..... 157–179  
         Gleevec ..... 161  
         vemurafenib ..... 132, 179  
     Mps1 ..... 176  
     p38 ..... 35  
     PDK1 ..... 178  
     PIM1 ..... 132, 167, 176

protein ..... 157–159  
 protein kinase B ..... 89, 90, 93, 95  
 Rho ..... 174, 177  
 serine ..... 90  
 TGF-beta ..... 140, 141  
 threonine ..... 90  
 tyrosine receptor ..... 158  
 Kinome ..... 170

**L**

LE. *See* Ligand efficiency (LE)  
 LEI. *See* Ligand efficiency indices (LEI)  
 Library design ..... 43–52, 121, 171, 175  
 Ligand efficiency (LE) ..... 78, 90, 121, 127, 163, 165, 179  
 Ligand efficiency indices (LEI) ..... 89–98  
 Ligand grid free energy (LGFE) ..... 82  
 Ligand lipophilicity efficiency (LLE) ..... 121, 163  
 Ligand–receptor complex ..... 4, 23–29, 124, 191, 192  
 Lipophilicity ..... 122, 124, 175  
 $\log P$  ..... 122, 163  
 LUDI ..... 60, 65–67, 69, 122, 123, 129, 168  
 Lysozyme ..... 8, 148, 193

**M**

MacroModel ..... 37–39  
 MarvinSketch ..... 103, 107, 110  
 Mass spectroscopy ..... 137  
 MCSS. *See* Multiple copy simultaneous search (MCSS)  
 MD. *See* Molecular dynamics (MD)  
 Metabase ..... 44  
 Metasite ..... 60, 68  
 MM. *See* Molecular mechanics (MM)  
 Molecular dynamics (MD) ..... 5, 8, 10, 11, 68, 76, 77, 79, 84, 102, 122, 123, 148, 149, 151  
 Molecular interaction field (MIF) ..... 63, 126  
 Molecular mechanics (MM) ..... 6, 10, 11, 14, 16, 19, 33, 37, 38, 95, 123  
 Molecular mechanics-generalized Born/surface area (MM-GB/SA) ..... 6, 15, 17, 19  
 Molecular mechanics-Poisson–Boltzmann/surface area (MM-PB/SA) ..... 6  
 Monte Carlo  
   Grand canonical Monte Carlo  
 (GCMC) ..... 33, 139, 147, 152, 167, 193  
   with variable chemical potential (SACP) ..... 193  
 Multiple copy simultaneous search (MCSS) ..... 60–63, 68, 122, 123, 167, 168  
 Multiple quantum number (MQN) ..... 122  
 Multiple solvent crystal structures (MSCS) ..... 146, 147  
 Multi-scale Truncated Newton (MSTN) ..... 17

**N**

Neuronal nitric oxide synthase (nNOS) ..... 59, 62, 65  
 New chemical entity (NCE) ..... 137  
 NMR. *See* Nuclear magnetic resonance (NMR)  
 nNOS. *See* Neuronal nitric oxide synthase (nNOS)  
 Normal mode analysis (NMA) ..... 148, 149, 151  
 Nuclear magnetic resonance (NMR) ..... 14, 19, 43, 57, 68, 76, 78, 102, 119, 120, 137, 138, 148, 151, 157, 161, 164, 166, 167

**O**

OPLS force field ..... 17

**P**

Pan-assay interference compounds (PAINS) ..... 51, 120  
 Partial charge  
   calculation of ..... 11, 33  
   by AmberTools ..... 103, 104  
   by Gaussian ..... 103, 104  
   Gasteiger–Marsili ..... 67  
   Momany–Rone ..... 63  
   RESP ..... 104, 107, 110  
 PDBbind ..... 102  
 PDB2PQR ..... 96, 102, 104  
 Pharmacophore  
   modeling programs ..... 68  
     Catalyst ..... 60, 68  
     GALAHAD ..... 60, 68  
     LigandScout ..... 60, 69  
     MOE ..... 69  
     phase ..... 60, 68  
 PocketFinder ..... 146  
 Pocket similarity ..... 23–29, 124  
 Polarized protein-specific charge model (PPC) ..... 10, 11  
 Polar surface area (PSA) ..... 34, 67, 90, 93, 121, 124, 132  
 PPI. *See* Protein–protein interaction (PPI)  
 Principle component analysis (PCA) ..... 63, 64  
   CPCA ..... 63, 64  
 Privileged structure ..... 124, 166  
 Proliferation assay ..... 202–204  
 Protease ..... 8, 9, 79, 132, 133, 147, 148  
   human immunodeficiency virus (HIV) ..... 8, 9, 132, 147, 148  
 Protein Data Bank (PDB) ..... 9–11, 14, 27, 61, 82, 91–94, 102, 104, 128, 132, 140, 141, 159, 160, 162, 168, 171, 172, 174, 175, 215, 217, 219, 222–224  
 Protein–ligand interaction ..... 6, 14, 35, 63, 101, 102, 147  
 Protein Local Optimization Program (PLOP) ..... 15–17  
 Protein–protein interaction (PPI) ..... 18, 59, 64, 68, 120, 145, 147, 148, 150  
 PSA. *See* Polar surface area (PSA)

## Q

- Quantitative structure-activity relationship (QSAR).....128, 131
- 3D QSAR .....69, 128

## R

- Rapid elimination of swill (REOS) .....124
- RECAP .....124, 126, 129, 168, 169
- Recore .....123, 124, 127, 129, 168
- Rule of 3 (Ro3).....51, 121
- Rule of 5 (Ro5).....121, 163

## S

- SAR. *See* Structure activity relationship (SAR)
- SBDD. *See* Structure based drug design (SBDD)
- Scaffold.....32, 58, 59, 69, 90, 92, 96, 121, 123, 125–132, 142, 143, 151, 153, 166, 168, 170–172, 174–177, 179, 190, 191, 219, 222–224 hopping .....58, 59, 126, 129, 142, 143
- Scanning electron microscopy (SEM) .....204–206
- Scoring function
  - ASP .....67
  - ChemPLP .....67
  - ChemScore .....67, 91, 95
  - CScore .....67
  - D\_Score .....67
  - G\_Score .....67
  - HYDE .....91, 96
  - MM-ISMSA .....91, 95
  - PMF\_Score .....67
- Seed .....60, 69, 84, 105, 108, 112
- SEI. *See* Surface efficiency index (SEI)
- SILCS. *See* Site identification by ligand competitive saturation (SILCS)
- SIMPLEX algorithm .....98
- SiteFinder .....146
- Site identification by ligand competitive saturation (SILCS) .....75–84, 123
- SiteMap .....60, 61, 146
- SMARTS .....45
- SMILES .....44, 46–48, 94

- Solvent accessible surface area (SASA) .....64
- SOLVMAP .....17
- SPAM .....7–10
- SPR. *See* Surface plasmon resonance (SPR)
- Structure activity relationship (SAR) .....5, 7, 8, 32, 69, 102, 119, 128, 143, 152, 161, 163, 167
- Structure based drug design (SBDD) .....23, 82, 101, 102, 132, 137–139, 145, 153, 174, 175
- Sub-graph mining .....28
- Subset .....15–17, 102, 122, 131, 132, 188, 200, 214
- Substrate activity screening (SAS) .....68
- Surface efficiency index (SEI) .....90, 91, 93–95
- Surface plasmon resonance (SPR) .....43, 57, 68, 76, 120, 137, 164, 174, 176

## T

- Tanimoto coefficient .....18, 122
- Tanimoto similarity .....47, 52
- Tethering .....57, 68, 128, 138, 175, 178
- Thermodynamic integration (TI) .....5, 138
- Tipranavir .....132
- Toxicophore .....61, 65, 124, 131, 132
- Transferable interaction potential (TIP) .....4

## V

- Virtual fragment .....31–39, 126, 153, 157 linking .....51
- Virtual library .....23–29, 121, 124
- Virtual screen .....11, 14, 92, 120–122, 125, 139–141

## W

- Watermap .....5, 8

## X

- X-ray crystallography .....34, 43, 57, 68, 76, 78, 119, 120, 123, 137, 151, 158, 164, 165, 167, 170, 172, 173, 175–178

## Z

- ZINC .....15–18, 126