Library Design, Search Methods, and Applications of Fragment-Based Drug Design
Library Design, Search Methods, and Applications of Fragment-Based Drug Design
Library Design, Search Methods, and Applications of Fragment-Based Drug Design

Rachelle J. Bienstock, Editor
National Institute of Environmental Health Sciences
National Institutes of Health
Research Triangle Park, North Carolina

Sponsored by the
ACS Division of Chemical Information
ACS Division of Computers in Chemistry

American Chemical Society, Washington, DC
Distributed in print by Oxford University Press, Inc.
Foreword

The ACS Symposium Series was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of the series is to publish timely, comprehensive books developed from the ACS sponsored symposia based on current scientific research. Occasionally, books are developed from symposia sponsored by other organizations when the topic is of keen interest to the chemistry audience.

Before agreeing to publish a book, the proposed table of contents is reviewed for appropriate and comprehensive coverage and for interest to the audience. Some papers may be excluded to better focus the book; others may be added to provide comprehensiveness. When appropriate, overview or introductory chapters are added. Drafts of chapters are peer-reviewed prior to final acceptance or rejection, and manuscripts are prepared in camera-ready format.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previous published papers are not accepted.

ACS Books Department
Preface

This volume is the result of Division of Chemical Information (CINF) symposia organized around the popular topic of fragment-based ligand design conducted during two Spring American Chemical Society National Meetings. The first symposium, Library Design, Search Methods and Applications of Fragment-Based Drug Design, was held during the 2009 Spring National ACS meeting in Salt Lake City, Utah. The second Fragment-Based Drug Design: Success Stories Due to Novel Computational Methods Applications, followed one year later at the 2010 Spring National meeting in San Francisco. All presenters at both meetings were invited to submit a written chapter summary of their oral presentations for this volume. The sessions at both of these meetings were extremely popular and the talks were presented to standing room only audiences. This reflects the strong interest in the technique and application of fragment-based drug discovery computational associated methods.

Computational, ‘rational’, or structure-based drug design methods have been promulgated since the early 1980s evolving over time. From statistical methods like QSAR, to shape based methods such as pharmacophore analysis, combinatorial chemistry and high throughput virtual screening, different computational methods have become part of the pharmaceutical company drug discovery arsenal. As pharmaceutical research and development time and costs have increased, there is constant commercial pressure to develop new and better methods for the identification of novel chemical entities. The success of the human genome project followed by protein structure initiatives, and molecular biology pathway analysis, have brought forth a plethora of new ideas for understanding disease pathways and new targets ripe for drug discovery. Into this arena has stepped the methodology of fragment-based drug or ligand design. This volume covers computational methods in fragment-based ligand design and their application to fragment library design, library screening and fragment docking methods and computational methodologies for fragment linking, merging and growing. It touches on some success stories in the development of potential leads using these fragment-based computational methods as well.

Readers who are new to the field of fragment-based ligand design as well as veterans of computational drug discovery methods and pharmaceutical and medical chemistry will be interested in the contents of this volume. In addition to chemists, mathematicians and statisticians, may be interested in this volume as well, in that they may see where novel algorithms, mathematical and statistical techniques, can be applied to the difficulties of library searching, and for the development of better fragment docking and scoring methods. Fragment-based drug discovery has rapidly risen as evidenced by both the increasing number of
industrial and academic groups with interest and publications in this area as well as documentation of at least 13 different companies (Abbott, Astex Therapeutics, Aventis, Burnham Institute, Novartis, Plexxikon, Roche, SGX Pharmaceuticals, Schering-Plough, Sunesis, Triad, Vernalis, Vertex) with successful leads as a result of fragment-based screening (Hajduk PJ and Greer, J Nature Reviews Drug Discovery 6, 2007, 211-219.) The development of twenty-two clinical candidates currently in Phase I, or II clinical trials are attributed to the application of fragment-based drug discovery methods (Law, R J Comput Aided Mol Design 2009, 23:459-473).

I would like to thank all those who agreed to write chapters for this volume as well as all those who participated in the two ACS symposia on this topic. Both symposia were filled with interesting presentations and discussions. I would also like to thank, Mr. Tim Marney, my editor at ACS publications for his assistance with all the intricate technical details involved in producing this volume, as well as Mr. Bob Hauserman, Senior Acquisitions Editor at ACS for noting the topical significance of the symposium and its merit for book publication. Of course, I always owe a debt of gratitude to my family, my husband Scott Snyder and daughters, Julia and Shira, for the support which they provide throughout all life’s endeavors.

Dr. Rachelle J. Bienstock

Contract Senior Research Scientist
National Institute of Environmental Health Sciences
National Institutes of Health
P.O. Box 12233, MD F0-011
Research Triangle Park, North Carolina 27709
(919)541-3397 (telephone)
biensto1@niehs.nih.gov (e-mail)
Chapter 1

Overview: Fragment-Based Drug Design

Rachelle J. Bienstock
National Institute of Environmental Health Sciences, P.O. Box 12233, MD F0-011, Research Triangle Park, North Carolina 27709

Fragment-based drug design has recently risen to great prominence as a new methodology for novel lead identification. This chapter is a general overview of computational methods for all three phases of fragment-based ligand design: (1) Designing and Searching Fragment Libraries, (2) Computational Screening: Docking, and (3) Leads from Fragments: Fragment Growing and Linking. Appendix 1 at the end of this chapter summarizes a large number of computational methods and software programs available with associated web sites and references.

Introduction

Fragment-based drug discovery has emerged as a new and promising computational methodology for efficiently increasing diversity space leading to novel leads and therefore NCEs (new chemical entities). For many researchers in the drug discovery area, high throughput screening (HTS) has not lived up to its original “great expectations”. Many HTS screens failed to deliver good starting molecules for drug discovery, as often the databases searched were composed of nondrug like molecules or compounds which, as products of combinatorial approaches, were difficult to transform into lead compounds. The preference for fragment-based design over other methods, such as high throughput screening, rests largely with the enhanced screening of a more impressive conformational space with a smaller starting number of compounds. A large HTS screen library contains $10^5$-$10^6$ compounds, which still samples only a fraction of the chemical space of small molecules (on the order of $10^{60}$-$10^{100}$ compounds), but world require huge amounts of testing. However, combinatorial combination of 3 different fragments of a small 100 fragment database would yield $10^6$ different compounds and require less experimental assaying or virtual screening (1).
Significantly higher hit rates have been reported with fragments than with HTS. HTS has on the order of a .1% or less successful hit rate while fragment-based design is estimated to have typically a 3-5% hit rate (depending on the target, hit rates are typically 8-10% for kinase targets and 2-3% for protein-protein interactions), and because the fragments are small when docked into the binding site, the binding site can be explored in new ways not accessible to large molecules (2). With fragments there is higher probability that a designed compound can be synthesized because it will be based on a small simple scaffold. In short, fragment-based design marries elements of rational design with the diversity of random virtual screening which makes it particularly attractive.

What are fragments? Fragments are low molecular weight (MW< 250Da), small organic molecules, that actually have low affinity (100 μM–10 mM) for binding to the target. These fragments are then embellished, grown and linked to create high affinity lead compounds with high selectivity. However, because these smaller fragments are weaker binders than leads and hits identified through HTS, the experimental methods for confirming binding must be more sensitive. Hits from fragment screening methods usually would not be identified as potential leads in HTS screens, and therefore provide for novel templates.

How do you design a good, diverse fragment library or database? Usually 2D fingerprint methods (Tanimoto and/or Daylight are among some popular methods) or statistical models based on fingerprint connectivity or pharmacophore models are used to analyze the library database for diversity (3). A popular set of rules, referred to as the “rule of three” has been proposed for the fragments comprising a library: MW less than 300 Da; less than or equal to 3 H bond acceptors or donors; ClogP less than or equal to 3; 3 or less rotatable bonds and a polar surface area less than 60 Å². “Drug-likeness” is an important characteristic for a fragment library and frequently fragment databases are compared to a database of known drug compounds, or often known drugs are dissected into fragments to develop a fragment database. High ligand efficiency (LE> or = 3.0), defined as the binding energy per heavy atom of the structure, is required for good leads.

Once a fragment library has been developed, a screening method must be implemented where fragments are screened against the pharmaceutical target of interest. Screening can be performed virtually in silico using computational methods, or experimentally using x-ray, NMR(SAR by NMR) (4), surface plasmon resonance (SPR, Biacore), mass spectrometry, isothermal titration calorimetry or protein thermal unfolding experimental methods. Soaking crystals with small fragments can be used as a method to identify fragments as well, the CrystaLEAD method of x-ray based fragment library screening (5).

Computational screening, using docking methodologies for fragments, is particularly challenging due to some of the same concerns as general docking. Problems with docking fragments include identification of the interaction or binding site for the fragment; binding cavities can be much larger than smaller fragments so there can be difficulties in predicting binding modes, and scoring functions are not optimized for fragments as they are designed for larger drug-like molecules. For fragments a metric other than RMSD must be used when docking, such as a structural fingerprint score, feature scoring, volume clustering, or fragment pharmacophore feature identification (6). In a review by Marcou and
Rognan of FlexX, Glide, Gold and Surflex on 42 protein–fragment complexes compared to x-ray data these docking algorithms could predict correct binding of fragments 40, 70, 70 and 60% of the time respectively (7).

Difficult targets for which fragment-based screening is particularly beneficial are those that are large and open to solvent and for targets which do not have well developed compound libraries. Structural information, such as from x-ray or NMR studies of the target must be available for both computational docking of fragments and successful linking and growing of fragments to fill the target binding pocket. As a result, many of the first successful applications of fragment-based ligand design have been exclusively in the oncology therapeutic area with many focused solely on kinase targets (8). Fragments are in general more rigid than molecules and easier to dock computationally due to fewer degrees of freedom. Fragment drug design is based on two fundamental assumptions- molecular recognition by receptor occurs due to the presence of a fundamental core structure or fragment, and that the properties of active fragments are additive and can be combined.

Fragment approaches can be complimentary to other lead generation methods and are often used in conjunction with other methods. Sometimes a set of fragments is assayed at high concentration (1mM) simultaneously with typical fragment screening concentrations (100 μM). Also sequential methodologies are employed following fragment-based design with a more sensitive screening technique. Often different hit finding methods work better with different types of targets. For example for the BACE-1 Alzheimer’s target, a high throughput μM screening of 200,000 compounds at Evotec did not yield any leads, however a screen of a 20,000 compound fragment library led to two dozen confirmed good hits tested as 1mM binders using SPR (surface plasmon resonance) testing. In fragment-based drug design frequently the idea of a “privileged structure” is employed where certain types of substructures are effective with particular types of targets, i.e. hydrosamates with matrix metalloproteases, benzamidines with serine proteases, aminopyrididines with kinases and ATP containing proteins (9).

Fragment-based discovery has claimed early successes with 50 perspective small molecule leads with good ligand efficiency (10). Several companies are developing drugs based on fragment screening which currently are in the clinic, including Abbott, Astex, SGX Pharm, and Plexxikon. Recently, on the practical fragments blog (http://practicalfragments.blogspot.com/) edited by Drs. Dan Erlandson and Teddy Zartler a list of 44 companies were compiled with research efforts in the area of fragment-based ligand design and drug discovery methods. These are examples of some targets where fragment-based ligand design has been successfully applied: BACE-1 (beta-secretase) for Alzheimers disease (11–13), Urokinase (14), Phosphodiesterase 4 (15), Bcl-XL for cancer,( a protein-protein interaction target) (16); Thrombin (17) Aurora kinase inhibitor (18) HSP90 inhibitor (19).

This volume covers the development of computational methods and their application in the area of fragment-based drug discovery. There are three basic steps involved in fragment-based computational drug design. The initial step is the design of a good fragment library, the second step is computational docking, ranking or screening of the fragments within the library and the third step is computational methods for growing, linking or combining of the fragments to
yield lead compounds. In the application of computational in silico screening
methods for fragments, many of the same issues and concerns apply as do
for computational docking of compounds in general. The targets used in
computational screening, usually structures solved by NMR or x-ray methods, are
rigid without flexibility in computational docking, and must be properly prepared
adding hydrogens, and side chains assigned proper protonation states. There is
a need for improved computational methodologies to score, rank and categorize
fragments docked to targets as well as for methodologies for growing or linking
the fragments to form complete molecules.

All three computational aspects were discussed within these ACS symposiums
and I will briefly outline the discussions and presentations in these areas. The
chapters within this volume discuss each of the methods presented, described by
their developers. Successful applications to kinase, GPCR, CNS targets, HSP 90
and drugs targeting protein-protein interfaces were presented at the meeting. As
we are focusing solely on computational methods in this volume, experimental
methodologies will not be discussed.

I. Designing and Searching Fragment Libraries

What comprises a good fragment library? The properties of a good fragment
library are : diversity of physicochemical properties, molecular diversity, aqueous
solubility, drug-like molecules, MW < 300 Da, good ligand efficiency (LE) (free
energy of binding a ligand averaged over each non hydrogen atom) and involve
taking lipophilicity into consideration. Several novel computational approaches to
the development of fragment libraries were reported at this ACS symposium. This
is a rapidly developing area involving both chemoinformatics and modeling tools.

One approach to the design of fragment libraries is the use of methods which
perform computational deconstruction of known drugs (20) for example the
DAIM method developed by Peter Kolb and Amedeo Caflish (Decomposition and
Identification of Molecules) (21) or retrosynthesis and combinatorial analysis,
such as the RECAP method (22). CoLibri is a commercial program (BioSolveIT)
that can be used for “shredding” compounds for the development of fragment
libraries for virtual screening.

The FTrees-FS software (23) performs fragment space similarity searching
and fragment assembly. (Searching web interface freely available: http://
public.zbh.uni-hamburg.de/ftrees/query.py; software available commercially from
BioSolveIT). Dr. Carsten Detering (BioSolveIT) reported on the development
and application of the FTrees-FS methodology and its advantages in using
synthetically accessible compounds from a giant virtual chemistry fragment space
library. FTrees-FS represents molecules in reduced graph representation (feature
trees) for easier and faster searching and extracts cores and identifies link atoms
,with a reagent list, to attach link atoms (24). Information concerning chirality and
3D properties of molecules is not included, only topology information
(25)(26)(27)(28). Dr. Atipat Rojnuckarin, reported on work conducted at ArQule,
implementing a novel targeting strategy to design type IV kinase inhibitors. This
involved application of the FTrees-FS software to construct searchable fragment
space based on the ArQule kinase inhibitor compound library to identify novel type IV kinase inhibitors with improved ease in synthesis.

Dr. J. Robert Fischer, Zentrum für Bioinformatik, Universität Hamburg, described LoFT, a library optimizer using Feature Trees, which is a tool for focused combinatorial library design (29). LoFT uses a reduced topological graph descriptor to match feature tree nodes to compare and search fragment space. Using LoFT, a fragment space can be searched and cores and reagents selected according to selected physicochemical properties. LoFT will then optimize and compile a complete fragment sublibrary with the described properties based on filters and descriptors in the scoring function. LoFT searches a fragment space consisting of combinatorial libraries with a unique scaffold and uses feature tree descriptors representing the molecules as unrooted nodes on a tree with topology, connectivity and physicochemical properties conserved. The difference between FTrees-FS and LoFT, is that LoFT results in sublibraries which focus on a single core placement for a more focused library design. For validation, LoFT was applied to several drug design scenarios. Starting with known drug molecules, focused libraries were generated with desired property ranges.

Dr. Christof Wegscheid-Gerlach, Bayer-Schering Pharma, reported on applications of BRICS (Breaking into Restrosynthetically interesting Chemical Substructures), a modification of RECAP to fragment molecules according to 11 default bond rules including comprehensive modeling of ring substitution and cleavage of sulfur groups. (BRICS fragment spaces are publicly available-http://ww.zbh.uni-hamburg.de/BRICS) (30) BRICS is a compilation of a new rule set for breaking up interesting chemical structures into fragments. Dr. Wegscheid-Gerlach discussed applications of BRICS to shred the WDI and Zinc databases and compared BRICS enhanced performance compared to RECAP. He then presented several successful applications demonstrating novel scaffold hops yielding Sorafenib, Fasudil and Erlotinib.

BROOD (Commercial software from Openeye http://www.eyesopen.com/brood) searches databases of chemical fragments to identify and select fragments with similarities to the query fragment and can perform bioisosteric replacements to develop new leads. BROOD also generates analogs to leads by assembling and replacing different fragments based on shape, electrostatics and molecular properties with graphical tools for fragment editing. BROOD is accompanied by CHOMP which serves as a molecular fragmentor and MERGE for fragment merging. Additionally, if a crystal structure of the target protein is known, BROOD can use information from the protein structure to eliminate fragments which will not fit in the binding pocket and will verify protein-ligand close contacts (31).

Dr. Ijen Chen, Vernalis, reported on the use of SeeDs (Structural Exploitation of Experimental Drug Startpoints) a method which uses pharmacophore fingerprints to facilitate fragment library design. The SeeDs library enumeration creates diversity through the use of pharmacophore triangles to create a fragment library. Fingerprints of 3 point pharmacophore triangles (acceptor, donor and hydrophobid) are used and these pharmacophore fingerprints can screen drug like chemical space and pick out the known preferred substructures (32, 33). The first SeeDs library was based on MW and desired chemical features judged
by medicinal chemistry, solubility and tractability. This was used for in house docking followed by a merging of fragments to create novel leads for PDK1 and Hsp90, ATPases and kinases. Average screen hit rates of 5.6% were reported by Dr. Chen for the SeeDs library screen on a dozen diverse targets.

Dr. Francois Delfaud (MEDIT-SA http://medit-pharma.com/index.php?page=MEDIT-SA-Products-Drug-Design-Softwares) reported on Med Sumo, a database searching method to identify similar binding surfaces of macromolecules based on specific similar chemical features and properties. Med Sumo can be used in conjunction with Med-Portions to identify small ligand fragment structures binding to these surfaces. Med-Portions is a protein-ligand database which includes a binding site database and protein-fragment database. Once the Protein-Ligand has a solved structure in the PDB, it can be converted to a protein-fragment pattern (MED-Portions). The Med-Portion database can be mined with a library of small molecules for detection of protein pocket similarities and alignments to identify new binding fragments. It generates fragments sharing some surface interaction features with the query protein-ligand solved structure surface (taken from PDB). A chapter in this volume describes this work in detail.

Dr. Valerie J. Gillet and her colleagues in the Chemoinformatics Research Group, University of Sheffield, reported on a de novo design method using reaction vectors and its application to fragment library design. In fragment library de novo design, one of the concerns is the design of molecules that can be synthesized easily and cost effectively. Dr. Gillet’s method for library design involves a knowledge-based approach using reaction vectors that describe structural changes at the reaction center within a reaction database. Dr. Gillet has written a chapter for this volume which describes her method.

Dr. Qiong Yuan and her group at Chemical Abstracts Service described SubScape for SciFinder (www.cas.org/products/scifindr/subscape) and its use for substructure searching to facilitate FBDD (fragment-based drug design). Fragments can be analyzed and sorted and organized with associated experimental and predicted properties with each substructure including bioactivity. Chemical space around fragments hits can be optimized using 2D Tanimoto similarity searches (www.cas.org/products/scifindr/subscape).

Dr. John Badger, DeltaG Technologies in conjunction with Zenobia Therapeutics, reported on the design and application of fragment libraries for crystallography studies. Zenobia applied this strategy to a CNS target (Parkinson’s disease), LRRK2, using rule-based filtering software to generate appropriate fragments for crystallographic screening methods. Compounds were searched for drug like core substructures (SDSearch) and then the rule of three was implemented and additional special useful filters (i.e. blood-brain barrier permeability) were used (34). SD search is an in house search tool which is comprised of Zenobia’s small fragment library. The Target LRRK2- leucine rich repeat kinase was use for “pseudo docking” the lead fragment. In this way key binding interactions could be indentified and a first scaffold screen with a focused library was performed followed by a second round focused around the best hits. LeadModel3D was used for docking. The best hits were selected for an experimental screen. Dr. Badger describes this work in detail in a chapter in this volume.
Drs. Ammar Abdo, and Naomie Salim, Universiti Teknologi Malaysia, Faculty of Computer Science & Information Systems, introduced a novel similarity-based virtual screening approach based on a Bayesian interference network. Their network permits a combination of multiple queries and molecular representations and weighing schemes. They feel that their method surpasses the Tanimoto similarity approach and offers a reasonable method to assess 2D similarity between structures. A chapter in this volume discusses this work.

Dr. Tobias Lippert, Zentrum für Bioinformatik, Universität Hamburg, presented QSearch: a pharmacophore-based search in fragment space. QSearch is an iterative search method using molecular evolution to allow a search of fragment space for molecules that can fulfill the criteria of a three dimensional pharmacophore. The search method uses an evolutionary approach where partial solutions evolve to fit the posed query by adding, deleting or replacing fragments. The fitness of a partial solution is calculated by its ability to obey the constraints of the pharmacophore. An example was presented using a thrombin query (PDB structure 1c4v) and focused fragment space as input with known thrombin inhibitors cleaved with BRICS rules which gave 800 fragments.

Dr. J. D. MacCuish and colleagues at Mesa Analytics & Computing, Inc., reported on a method for shape clustering of fragment databases using both 3D shape fingerprints (generated via Quasi-Monte Carlo integration) and 2D structure fingerprints. Individual clusters are then analyzed with 3D shape fingerprints incorporating substructure information, akin to substructure commonality programs with 2D fingerprints, such as Stigmata and ChemTattoo.

Dr. V. V. Poroikov, Department for Bioinformatics, Institute of Biomedical Chemistry of Rus. Acad. Med. Sci., reported on the PASS method that predicts more than 3000 biological activities of a database of chemical compounds (http://www.ibmc.msk.ru/PASS) (35). Prediction is based on SAR (structure activity) analysis of the training set containing over 200,000 biologically active compounds collected from different sources. PASS calculates the impact of each atom in a molecule into a certain activity using MNA descriptors for each particular atom and its immediate neighbors. These estimations could be used for identification of fragments responsible for binding chemical compounds with a specific target, and for further computer-aided design or generation of new "candidates" with the required biological activity.

Drs. Y. Xu, H. Jansen, and E. Martin of the Novartis Institutes for Biomedical Research, proposed a method when binding site information is known, modifications can be proposed using a "cut and fit" and "fit and cut" method. The "cut and fit" approach fragments a compound database replacing part of a lead molecule with fragments; the fit and cut starts with a complete molecule from a compound database, and determines whether this molecule fits. The fragments selected are separated and merged with relevant parts of the lead molecule. Finally, the fitting of the new modifications are confirmed with docking method. This method has produced interesting ideas in multiple kinase projects. As a validation of the method, a case study with the P38 ATP pocket and the MDDR database was described.

ALTA (Anchor-Based Library Tailoring) a focused chemical library obtained by prioritizing molecular fragments according to their docking energy is a
technique developed by Dr. Peter Kolb and Dr. Caflish (36). An example was presented in the talk by Dr. Kolb of the small molecule ABT-737 mapping bound to Bcl-XL and this method is discussed further in a chapter by Dr. Kolb in this volume.

II. Computational Screening: Docking

After designing fragment libraries and screening these libraries for hits, the next step in the process is to computationally dock these fragments into the targets or receptors to determine energetically favorable binding site positions for fragments and functional groups. Once a good fragment hit is found, it is developed into a lead by linking, growing or merging. Fragments can be positioned in the binding cleft of protein targets and then grown, attached or linked to fill the binding pockets optimizing steric, electrostatic, van der Waals and hydrogen bonds. Some of the commonly used software methods available for fragment positioning in the past are the methods GRID, MCSS, SPROUT, MUSIC, LUDI, Sklegen (De Novo Pharmaceuticals) and Superstar (CCDC). At the ACS symposiums presentations were given on the methods whose descriptions follow.

The MCSS method (Multiple Copy Simultaneous Search), and the Miranker and Karplus paper (1991), are considered by many to be the one of the first examples of fragment-based ligand docking. The MCSS method (currently implemented as a commercial product in Accelrys Discovery Studio software suite) uses simultaneous molecular mechanics minimization of fragments in the active site using CHARMM force field and the fragments are ranked by their MCSS energy score. The MCSS method can be used to determine fragment binding modes. Dr. Jürgen Koska and colleagues from Accelrys and Pfizer presented a paper at this meeting docking small fragments using MCSS minimization. Accelrys has taken the original MCSS method and incorporated it into a fully automated Pipeline Pilot workflow and demonstrated performance with scoring and placing of fragments with correct poses in several protein-fragment complexes.

Drs. Dina Kozakov, and Sandor Vajda, Boston University, gave a presentation on FTMAP, a method also based on MCSS, to find druggable sites at protein-protein interfaces using computational fragment mapping. Developing drugs which target protein-protein interfaces has been a recent area of significant interest. The FTMAP method uses small molecules, fragments or groups on the surface as probes and finds and clusters their most energetically favorable positions. In this way, the hot spots for drug binding are identified. When proteins interact as binding partners, although there might be a large binding surface area, there are specific essential “hot spots” where they interact. Using a small organic molecule probe, and an efficient FT algorithm (FTMAP) for sampling the surface area in a grid like manner, clusters of probe consensus binding sites can be identified. The largest consensus cluster is the most significant “hotspot” for binding. This methodology grew out of a experimental techniques for solving protein x-ray structures called Multiple solvent crystal structures (MSCS)
(37), where each solved structure is solved with multiple organic solvents. When these structures are compared, the organic solvent molecules cluster in consensus sites that are significant functional hot spots. The idea was to develop a computational method which could identify these functional hot spots to use in place of experimental x-ray crystallography. The output from FTMAP is a PDB file and the 6 lowest energy cluster representations for each probe, the number of nonbonded interactions between the probes and residues, and the number of H bonds between probes and each residue. These docked organic fragments can then serve as a starting point for fragment-based drug design. This group has applied this method to several protein-protein interaction systems, including interleukin-2, Bcl-xL, MDM2, HPV-11 E2, ZipA, TNF-alpha, and NEMO. The FT map server is freely available for users http://ftmap.bu.edu (38).

Dr. Peter Kolb, (Professor Amedeo Caflisch’s group, University of Zurich), gave a presentation on the application of several of the methods for fragment design developed by this group: DAIM, SEED, FFLD and GANDI. The Caflisch group has designed several publicly available programs (http://www.biochem-caflisch.uzh.ch/download/) which work together for fragment ligand design. DAIM for decomposition of molecules into a fragment library: SEED (Solvation Energy for Exhaustive Docking) for docking fragments and FFLD for molecule docking based on docked fragment locations found using SEED. After automatically decomposing molecules in a library into fragments (DAIM) they can be docked and the docked fragments ranked (SEED). SEED docks the fragments and the favorable poses of anchor fragments are used for FFLD (Fast Flexible Ligand Docking) which docks the molecules designed, so the two programs SEED and FFLD are designed to work together. The docking program SEED docks and orients fragments into a binding pocket (39, 40), and the binding energy estimated. The free energy of binding can be calculated for multiple poses (LIECE (linear interaction energy with continuum electrostatic method).

GANDI is a Genetic Algorithm-based de novo design of inhibitors and is a program for fragment-based de novo ligand design (41). GANDI performs automatic design of molecules within known binding site structures. It includes a novel simultaneous energy minimization and a term forcing 3D similarity to known inhibitors or ligands through 3D overlap. GANDI’s fragment method joins pre-docked fragments with linkers, which are evaluated with a search algorithm. Dr. Kolb has written a chapter describing applications of these methods in detail.

Dr. Zsolt Zsoldos, SimBioSys Inc, (http://www.simbiosys.ca/) gave a presentation on the application of the fragment-based docking and linking engine of eHiTS. Any docking method used for molecules can be used for fragments as well. However, many of the conventional docking methods have problems with fragments largely due to the fact that the cavities are large and the fragments small and therefore the fragments are not sufficiently constrained for docking within the cavity. Since eHiTS works by breaking down larger ligands into small fragments and docking them independently and then reconnecting the fragment poses it has resulted in about a 0.5 Å RMSD small fragment pose prediction and is capable of linking the fragments without loss of information (42–44). Dr. Zsoldos has a chapter in this volume describing applications of eHiTS for fragment docking.
III. Leads from Fragments: Fragment Growing and Linking

The time required to develop good leads from fragment-based screens can be longer than other methods since fragments involve additional development work to evolve into leads. Good methods for growing and linking fragments can aid this process. There are several general methodologies for creating a lead compound from fragments: (1) linking two or more fragments that bind to different parts of the target pocket to create a lead with a chemical bridge compound or linker (2) fragment self binding- where two or more fragments bind or connect to each other due to chemistry without a linker (“click chemistry” is an example) (3) simply optimizing the fragment itself alone and essentially using the fragment itself as a lead (4) “fragment evolution” where functional groups which bind to the target binding partner are added to increase the fragments affinity . Commonly used computational methods for fragment linking in the past include: Caveat, HOOK, Recore, Allegrow, Confirm, MED-SuMo LEGEND (45) LUDI (46), GROWMOL (47) LigBuilder (48) SkelGen (49, 50) SMoG (51) LUDI, HOOK (52), PRO_LIGAND (53), LigBuilder, SPLICE/RACHEL (54), CAVEAT (55, 56), CLIX and LUDI GROW, LEGEND, Lore, GEminini, GRID and many similar methods place fragments on grid points in the active site and determine favorable interactions. HSITE maps hydrogen bonding regions of the enzyme active site. Other methods for in situ fragment linking involve dynamic combinatorial chemistry, “click chemistry”, and fragment tethering- disulfide bond formation between cys residues in the protein and the fragment. 

GroupBuild (57) was one of the first design linking methods, described as a “fragment-based method”. GroupBuild uses a library of organic templates and a force field describing nonbond interactions between the ligand and enzyme to build drug candidates that have steric and electrostatic properties that fit with the enzyme binding site.

Dr. A. Peter Johnson and colleagues from the University of Leeds, reported on SPROUT (58), a computational tool for growing fragments (now available commercially through SimBiosys Inc. http://www.simbiosys.com/sprout/index.html). SPROUT was an older program originally developed for de novo ligand design, however it is useful for fragment linking as well. When information is known about a fragment and its binding pose, SPROUT with two or more fragments, is able to link them together, redocking to maintain the original poses, also permitting some movement, limited by user selected tolerances. SPROUT is also capable of fragment growth (evolution). SPROUT locates binding pockets and identifies potential interaction sites (H bonding, hydrophobic, covalent, metal, user defined) and docks molecular fragments to target sites, and generates novel chemical structures from templates and clusters. Using SynSPROUT, the fragments are linked through virtual synthetic chemistry. LeadOpt has reaction information and starting material information to predict virtual reaction and synthesis of the ligands which are known to bind and works together with SynSPROUT.

FlexNovo is useful for large fragment fragment space database docking (59), when the active site structure is known. FlexNovo is based on the FlexX flexible docking algorithm with structural information for the target active site
and pharmacophore type constraints for the fragments. The fragments in the library are combined and linked and docked in the active site with a calculated score (60). Recore, (http://www.zbh.uni-hamburg.de/en/research/computational-molecular-design/projects.html), (developed by Mattias Rarey,ZBH Hamburg, in conjunction with Hoffmann LaRoche AG Basel Switzerland, commercial product from BioSolveIt ) (61) is a new fragment replacement tool for fast searching of conformations (similar to the earlier CAVEAT) however combined with a large search domain focusing on druglike structures and including pharmacophore type searching. The result is that Recore provides for effective scaffold hopping, 3D core replacement and fragment linking and merging. The fragment database is created from 3D structures with cleavage rules defined by SMARTS patterns. While ReCore uses indexed searching, it additionally does core replacement, and fragment linking and growing and merging.

AlleGrow, is a program which can be used in the design of cyclic scaffolds which connect and incorporate fragments. It is an update of the earlier GrowMol (62)(RS Bohacek). Cyclic scaffolds are fairly common in bioactive molecules. CONFIRM (Connecting Fragments Found in Receptor Molecules) is a linking approach with a search library for bridges for fragments and automation of linking and docking to a target (63). A prepared library, of bridges and links, is used as linkers to fragments and docked. CONFIRM retrieves molecular fragments based on distances and atom types and searches a database of bridges using a substructure pattern search. These bridges are linked to the fragments in combinatorial ways and then proposed molecules are docked computationally.

Dr. Jacob Durant (Dr. McCammon’s group, University of San Diego) gave a presentation on AutoGrow a fragment growing method using an initial core and randomly adding fragments to the core scaffold which are then dynamically docked into the protein. It is a genetic algorithm so the compounds that dock the best become the scaffolds for the next generation in an iterative fashion (64). It is freely available for download: http://autogrow.ucsd.edu/. AutoGrow and AutoClick are based on an evolutionary algorithm which starts with an initial scaffold mutation operator and replaces it with a molecular fragment and docks it into the target structure.

BREED (licensed from Vertex and implemented as scripts fragment_join.py and fragment_link.py in Glide XP for fragment docking within the Schrodinger software suite http://www.schrodinger.com/scriptcenter/#/Fragments, and also available as part of the Chemical Computing Group MOE software suite http://www.chemcomp.com/software-sbd.htm) (65), is an automated computational method to create new inhibitors by joining fragments from ligands whose structures bound to the target are already known. Scaffold structures of known ligands are superimposed so that similar bonds match and can then be split and recombined in different combinations to generate new ligands. This method is dependent on having structural information. It is based on and similar to the earlier SPLICE method (66). BREED has been used successfully by Vertex to design HIV protease inhibitors and kinase inhibitors. Methods like BREED, that replace groups hanging off a scaffold belong to the group of methods referred to as “scaffold hoping”. BREED detects bonds from different ligands in close proximity in spatial alignment when the ligands are superimposed. Fragments
can be recombined to form chimeric compounds. This is referred to as “fragment shuffling” (BREED, RECORE, FLUX and MED-Hybridise are all methods in this category). In addition to BREED, Chemical Computing Group also has tools within the MOE software suite for scaffold replacement, fragment linking and growing, and medchem transformations including pharmacophore features (see http://www.chemcomp.com/journal/newscaffold.htm).

Dr. Peter Kutchukian (Harvard) gave a presentation on FOG, (Fragment Optimized Growth Algorithm), a statistically biased growth of fragments to produce compounds with certain features that appear with high occurrence in the training database (67). It uses a Markov Chain approach with branching treating each new fragment as a new transition probability and training on a database of bioactive compounds. Fragments must have shape and energetic complementary to the binding pockets, and synthetic feasibility so that certain types of molecular connections are favored in growing fragments and others forbidden. The result is the generation of molecules that have “druglikeness” properties (i.e. occupying drug like chemical space). Dr. Kutchukian has written a chapter in this volume describing FOG.

Dr. Dan Erlanson, (Carmot Therapeutics, Inc.), developed a technology, Chemotype Evolution (68), which uses rapid in-situ chemistry to expand a fragment into a diverse range of hits. Chemotype Evolution begins with a “Bait fragment” which is modified with a group or fragment and then used for screening so that a new “chemotype” is generated. This leads to custom library generation in an iterative fashion. The chemotype evolution method is directed at finding a “good fragment”, and is based on elaborating fragments found using any method: a starting “bait” fragment can be a “privileged” pharmacophore derived from a known inhibitor, substrate, or cofactor, or a fragment identified through a previous screen. Through iterative application of Chemotype Evolution, the starting fragment can be transformed into novel, varied “chemotype”, while desired properties can be enhanced by incorporating counter screens. This technique was applied to the challenging design of Aurora A fragments in an adaptive kinase pocket with DFG loop movement (69).

IV. Examples of Successful Applications of Fragment Design Process

Dr. Valerio Berdini, Astex Therapeutics, gave a presentation on the discovery of AT7519, a novel CDK inhibitor (which at the time of the meeting was in clinical trials) and AT9283, using fragment-based drug design methods. Fragments used for the CDK project were used to develop novel Aurora kinase inhibitors. This work led to the identification of AT9283 which is also was in clinical trials at the time of the meeting. AT7519 inhibited CDK2 with an IC50 = 0.047 micromolar and LE = 0.42 (70). AT9283 (Pyrazil-4-yl Urea) is an Aurora kinase inhibitor with IC50 = 0.91 micromolar and LE = 0.59 (71).

Dr. Miles Congreve, Heptares Therapeutics, discussed fragment-based screening of stabilized G protein-coupled receptors. GPCRs are difficult targets due to their conformational flexibility, heterogeneity and instability outside the cell.