In Silico Fragment-Based Drug Discovery: Setup and Validation of a Fragment-to-Lead Computational Protocol Using S4MPLE

Laurent Hoffer,[†,‡] Jean-Paul Renaud,[‡] and Dragos Horvath[†,*]

[†]Université de Strasbourg, 1 rue B. Pascal, Strasbourg 67000, France
[‡]NovAliX, BioParc, bld Sébastien Brant, BP 30170, Illkirch 67405 Cedex, France

S *Supporting Information*

ABSTRACT: This paper describes the use and validation of S4MPLE in Fragment-Based Drug Design (FBDD)—a strategy to build drug-like ligands starting from small compounds called fragments. S4MPLE is a conformational sampling tool based on a hybrid genetic algorithm that is able to simulate one (conformer enumeration) or more molecules (docking). The goal of the current paper is to show that due to the judicious design of genetic operators, S4MPLE may be used without any specific adaptation as an in silico FBDD tool. Such fragment-to-lead evolution involves either growing of one or linking of several fragment-like binder(s). The native ability to specifically "dock" a substructure that is covalently anchored to its target (here, some prepositioned fragment formally part of the binding site) enables it to act like dedicated de novo builders and differentiates it from most classical docking tools, which may only cope with non-covalent interactions. Besides, S4MPLE may address growing/linking scenarios involving protein site flexibility, and it might also suggest "growth" moves by bridging the ligand to the site via water-mediated interactions if H_2O molecules are simply appended to the input files. Therefore, the only development overhead required to build a virtual fragment→ligand growing/linking strategy based on S4MPLE were two chemoinformatics programs meant to provide a minimalistic management of the linker library. The first creates a duplicate-free library by fragmenting a compound database, whereas the second builds new compounds, attaching chemically compatible linkers to the starting fragments. S4MPLE is subsequently used to probe the optimal placement of the linkers within the binding site, with initial restraints on atoms from initial fragments, followed by an optimization of all kept poses after restraint removal. Ranking is mainly based on two criteria: force-field potential energy and RMSD shifts of the original fragment moieties. This strategy was applied to several examples from the FBDD literature with good results over several monitored criteria: ability to generate the optimized ligand (or close analogs), good ranking of analogs among decoy compounds, and accurate predictions of expected binding modes of reference ligands. Simulations included "classical" covalent growing/linking, more challenging ones involving binding site conformational changes, and growth with optional recognition of putatively favorable water-mediated interactions.
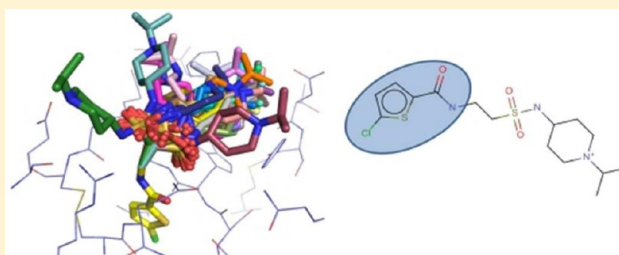
1. INTRODUCTION

1.1. Focus and Outline of this Work. This paper describes the use of S4MPLE for the in silico fragment-to-lead evolution step in Fragment-Based Drug Discovery (FBDD). This approach typically implies fragment concatenation operations, starting from experimentally validated bound fragment-like hits in the target site and has been addressed by several dedicated tools. It includes either linking of several independently binding fragments or growing of one fragment-like compound to lead-like size. For clarity purposes, the single term "linker" will refer to added atoms/groups with respect to starting hit(s) in both linking and growing contexts. In silico FBDD is basically a (structure-based) subdomain of de novo drug design (DND, see section 1.3), requiring a starting moiety of validated affinity and of known binding mode to the target.

The focus of this paper does therefore not lie on the generic stumbling blocks of FBDD/DND fragment-to-lead evolution: fragment library design, chemical space coverage, assessment of generated compound feasibility, etc. (see section 1.3). It is to show that, with a minimalistic set of tools for linker management

and focused library generation, both classical and more challenging in silico fragment-to-lead evolution can be successfully carried out with a generic conformational sampling program without further adaptation. In silico FBDD emerges, if one may say so, as a collateral benefit of the versatility of the operators designed within S4MPLE.

S4MPLE, a Lamarckian genetic algorithm for generic conformational sampling, has the particularity of full control over all the degrees of freedom (DoF). This is due to a deceptively simple and yet fundamental choice of generic operators recombining the local geometries of various molecular moieties. Rather than distinguishing, like many classical sampling/docking programs, between intra-molecular DoF associated to covalent bonds/torsional axes and intermolecular Euler angles/quaternions, S4MPLE completely ignores this arbitrary classification and proposes a moiety recombination scheme driven by favorable contact formations. The latter may be covalent bonds as much as

hydrogen bonds or hydrophobic contacts. It has been shown[1] how this choice allows S4MPLE to (A) simultaneously dock several entities, including free crystallographic waters added—or removed, when replaced by ligand groups—in order to predict multi-molecular complex geometries. Reversely, S4MPLE may (B) dock a piece of a molecule like a linker into a formal active site of the protein including the initial fragment-like hit(s) to which the linker is covalently connected. In addition, it may, if needed, also unlock DoF of specific protein moieties (linking/growing in flexible active sites). Eventually, it flawlessly supports a combined (A+B) scenario, trying to simultaneously include waters that might indirectly anchor the evolved lead to the site, i.e., formally "grow" the fragment by adding both covalent substituents and non-covalent partners (solvent molecules) mediating favorable interactions.

1.2. Short Overview of Fragment-Based Drug Design.
FBDD is an emerging technology that builds ligands from low complexity compounds called fragments.[2−4] Because of their small size, they often have low potency. Therefore, very sensitive methods, such as biophysical techniques (NMR, SPR, MS, etc.), are employed in the screening campaigns.[5−8] Besides, X-ray crystallography and NMR are both able to resolve the binding modes of fragments within the binding site—this information greatly facilitates the fragment-to-lead optimization using Structure-Based Drug Design (SBDD) strategies.[9] A main advantage of FBDD is that fragment hits often have strong affinity with respect to their pretty small size (i.e., good ligand efficiency LE[10]). Also, hit rates in fragment collections are higher than in drug-like ones.[11] Screening an appropriate diverse library (e.g., $10^4−10^5$ fragments[12]) is in principle equivalent to the exploration of a much larger drug-like space domain than accessible by traditional high-throughput screening.

There are three main strategies to optimize fragments into high affinity drug-like ligands: merging, linking, and growing.[13] Growing looks like usual hit2lead optimization because chemical groups are simply added to one single fragment. Both merging and linking strategies assemble two fragments that bind simultaneously. However, only linking adds additional atoms by means of a spacer. Ideally, binding modes of original fragments must be conserved in the final compound after merging/linking without adding undesirable strain energy, enabling the bonus gain in affinity associated with successful linking.[14] Linking is more challenging than growing, and there are some very interesting success stories where two low affinity fragments are linked into one high affinity ligand using a spacer.[15−17] However, it is not surprising that a lot of success stories in FBDD used growing optimization, with the support of X-ray data.[2]

1.3. In Silico Fragment-Based Drug Design.
In parallel to the growing interest in experimental FBDD, computer-based strategies, targeting main bottlenecks of FBDD (library design, fragment hits identification and their subsequent optimization into leads), are being developed.[18,19]

Chemoinformatics[20] can be used to design a fragment library with high scaffold diversity, while only incorporating compounds containing chemical functions known to be easily alterable in order to facilitate their subsequent growth.[21] When the 3D structure of the target is available, a prioritized fragment subset for experimental testing can be built by means of virtual screening approaches, such as docking or pharmacophore screening. Docking has been more or less successfully used to predict the binding mode of drug-like ligands,[22] and in practice, this technique can be applied to fragment-like compounds too. Small size implies low numbers of putative favorable contacts, thus

narrow binding energy ranges. While this challenge of scoring may become critical with very small compounds, a recent study[23] concluded that there is no significant difference in docking accuracies of drug-like ligands (sampling issues) vs fragment-like compounds (scoring issues). Good results were also reported with Glide[24] in both fragment redocking studies[25] and virtual screening.[26]

Finally, fragment→lead optimizations can be suggested by de novo design (DND) algorithms or similar in silico growing/linking strategies. The main goal of DND is to generate chemically sound and original structures predicted to have desired physicochemical and biological properties. The process can often be guided by using both structure-based docking, pharmacophore constraints,[27,28] or any other 2D/3D property prediction model.[29] Compound growth is based on chemical linkage rules, such as the RECAP[30] or BRICS[31] reaction schemes, in order to increase the synthetic accessibility of suggested compounds. LUDI,[32] Legend,[33] SPROUT,[34] SkelGen,[35] GroupBuild,[36] and LigBuilder[37] are common DND programs. TOPAS[38] and Flux[39] are RECAP-based DND software. For more details, see Schneider and Fechner.[40] Albeit a marginal topic in this work, it is important to highlight the importance of synthesis tractability of virtually generated compounds.[41−45] For example, generic reaction schemes[46] (reactants→products, with −R groups) are encoded in order to mimic real chemistry, as in the DOGS procedure.[42]

Various DND algorithms have been either explicitly developed to address some of FBDD issues or can be used in this context. The search for energetically favorable locations for various chemical entities is already known as probe mapping, and MCSS,[47,48] FTMap,[49−51] or GRID[52] are popular programs of that kind. After a search step that includes various organic probes (e.g., isopropanol, cyclohexane, benzene, etc,), FTMap performs a cluster analysis in order to identify putative hot spots. The latter are defined as consensus areas over several probes and obviously with favorable energies. Using the small but diverse "Congreve data set",[2] Hall et al. showed that their FTMap strategy was able to identify the main hot spots (defined as the subpocket that binds the studied fragment hit) as the top-ranked cluster, and some other clusters matched lead moieties evolved from these starting fragments.[50] Recore has been developed to perform scaffold hopping,[53] and relies on "exit vectors" defining moieties to replace. However, such vectors can be used to define the direction in which a compound must be optimized, hence the possible use of Recore within FBDD projects. Besides, pharmacophore and binding site constraints can be added to guide the evolution process. In the FBDD context, fragments linking algorithms are of particular interest when binding modes of simultaneously fragment binders are known. Two recent tools are specialized in the linking of predocked fragments: CONFIRM[54] and GANDI.[55] In CONFIRM, a pre-prepared library of "bridges" is screened to extract appropriate linkers in order to join predocked fragments. In this study, the Glide docking tool[24] was used to locate the starting fragments, and the validation of the method was performed on retrospective cases. The program GANDI uses fragments poses (e.g., generated by the SEED program,[56] developed in the same laboratory) as inputs too. It relies on a genetic algorithm involving an island paradigm and a tabu search during the sampling stage, and the energy function is based on the force field formalism.

To conclude, the three main issues of FBDD can be addressed, with more or less success, by computer-based approaches. Several reviews described both sides (experimental and computational) of fragment-based drug discovery.[4,57]

1.4. Introduction of the Virtual Screening Protocol.
In order to apply S4MPLE as a virtual screening engine for suited

linkers, the only additional developments required by this virtual fragment→ligand growing/linking procedure were two Java programs based on ChemAxon API:[58] a linker generator and a focused library builder, automatically connecting compatible linkers to the prepositioned reference fragments.

In a nutshell, S4MPLE may be employed to specifically optimize positions of linker atoms only, in search of the most stable linker conformation strainlessly connecting the prepositioned fragment(s)—one in growing and two in linking. Thus, the valuable information of the initial positioning of the starting fragment(s) can be exploited. S4MPLE will first optimize fragment-linker and intra-linker covalent structures under constraints given by the fixed ends of prepositioned fragments and the excluded volume of the (fully rigid or partially flexible) target site. While initial fragment(s) atoms do not move, but the site is likely to adapt its geometry to the linker, site-specific DoF can be enabled. Eventually, constraints are lifted and the so-far best poses of the now fully flexible ligand are allowed to freely relax. In order to identify putatively favorable water-mediated interactions during growing or linking, the user merely has to concatenate the desired number of water molecules at the end of the ligand candidate files produced by the library management tools.

This strategy (see Methods) is retrospectively applied to case studies from the FBDD literature. Its success is measured according to several criteria:

- ability to create the known reference ligands (and/or close analogs) from starting fragment hits
- ability to prioritize known actives relatively early within the ranked list of generated compounds (supposed inactive, although some of them can be real binders)
- ability to accurately reproduce the experimental binding mode of reference ligands
- ability to predict binding site conformational changes (when site flexibility is enabled)
- ability to produce a better ranking/prioritization of known actives involved in water-mediated interactions, upon explicit adding of water molecules to the simulation. Comparatively, the water-free growing procedure is expected to penalize the experimental ligand (due to unaccounted water-mediated hydrogen bonds).

2. METHODS

The virtual fragment→ligand growing/linking protocol is explained below. Eventually, data sets used in the retrospectives studies are introduced.

2.1. S4MPLE. S4MPLE (Sampler for Multiple Protein−Ligand Entities), a molecular modeling program based on a Lamarckian genetic algorithm, has been described previously.[1,59] This conformational tool, allowing the selection of considered degrees of freedom of the system, can be employed for a wide variety of simulation types: conformational sampling of ligands or small peptides and docking of both fragment-sized and drug-sized compounds. There is no explicit limit with respect to the number of considered entities—simultaneous docking of multiple ligands is supported. The energy function relies on the force field (FF) formalism and uses AMBER[60] and GAFF[61] to respectively simulate peptide and small organic moieties of the considered system. Here, all simulations are performed with the "Fit FF" energy scheme.[59] The control of conformational similarity is performed by a symmetry-compliant pair-based interaction fingerprint (PIF) that monitors two interaction types: close carbons contacts (based on C−C distance) and hydrogen bonds. Two configurations of the system are considered equivalent

if the Hamming[62] distance between their fingerprints is lower than a user-defined threshold. The program is written in object-Pascal and used in command-line mode.

2.2. Fragment-to-Lead Development Tools. Two Java programs, GenLinkersDB and JMolEvolve, based on the ChemAxon API,[58] serve to virtually evolve starting fragment(s) into lead-like or drug-like molecules. In other words, they are used in order to create a focalized library around one (growing) or two (linking) starting hit(s).

2.2.1. GenLinkersDB. GenLinkersDB creates linker libraries by fragmenting a database following RECAP cleavage rules. The RECAP algorithm[30] runs through molecular structures, detects bonds of interest using a predefined dictionary, and cleaves them while assigning specific flags to the resulting "loose" valences of atoms. These flags encode the chemical context or neighborhood of the atom in its molecule of provenience, much like FF atomic types. Once a molecule has been fragmented, it is possible to build it back by connecting fragments containing complementary flags. As for example, when an amide bond is broken, two fragments are created with two distinct but complementary amide flags (e.g., amide:1 for N−* and amide:2 for C(=O)−*). An extensive fragmentation (including single atom fragments, cuts between heteroatoms and cycles) is performed to maximize the number of generated linkers. Fragments with one or two flags are respectively saved in "growing" or "linking" databases. As already mentioned and for convenience, the term "linker" will refer to the added atoms/groups in both cases (growing or linking). A mass threshold of 300 Da is defined because the goal consists in optimizing precursor compounds into RoS-compliant[63] molecules. Raw fragment databases, immediately after the fragmentation step, are huge and contain many duplicates. These are removed in a two stages procedure:
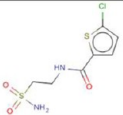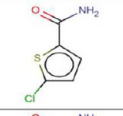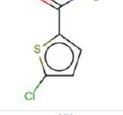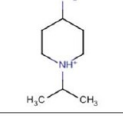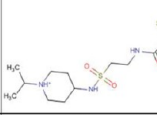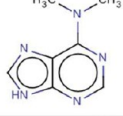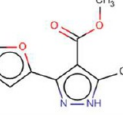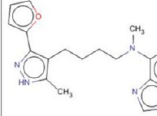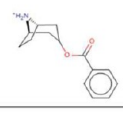
- Linkers are stored into different files, indexed by their mass (storage file M contains ligands of mass [M, M+1 (Da)].
- Specific functions (see below), based on ChemAxon graph isomorphism algorithms, are subsequently called to remove redundant entries within these smaller and more tractable files.

Indeed, linker uniqueness cannot be defined on the naked molecular graph but can be defined on the RECAP-flag property-labeled graph. A same skeleton may o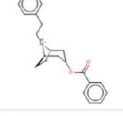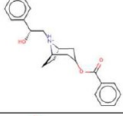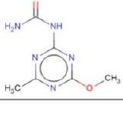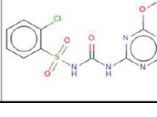ccur in conjunction to different or differently located RECAP flags. It should be noted that mass-splitting allows not only the acceleration of duplicate searches but also results in building complexity-ranked linker databases. The virtual evolution process can be started with the shortest linkers and stopped as soon as results show that desired linker size may have been passed.

The clean subset of the ZINC[64] version 12, containing more than 9 million molecules, is used as input file for the full fragmentation protocol (described above). The final numbers of linkers are about 2.2 and 3.6 million, respectively, for linking and growing databases. Subsets with lower mass limits or element constraints (e.g., only H, C, N, O, S, and P elements) are easily extractable therefrom.

2.2.2. JMolEvolve. JMolEvolve builds new molecules by merging one original precursor fragment and chemically compatible linkers based on the complementarity of the RECAP flags. Precursor fragments must be input as 3D molecule files with the coordinates corresponding to their docked or experimentally determined position relative to the active site and linkage points (e.g., fragment atom ID(s) to be used in linking/growing). This growing mode can be switched to linking mode by simply providing two precursors. In practice, the user must

Table 1. Description of All Growing/Linking Protocols[a]

Target	Protocol	Molecule ID	Structure	Binding data	RECAP flags	Linkers database	Filters	Generated & screened compounds	Binding site flexibility
FXa	Growing I (Small)	2		Ki ≈ 60 μM LE=0.36	sulphonamide	Growing Mass 250 Da HCNOSP	HAC < 12	≈ 4000 compounds Reference : yes Analogs : yes	Polar hydrogens
	Growing II (Large)	1		Ki ≈ 60 μM LE=0.58	amide	Growing Mass 250 Da HCNOSP	12 ≤ HAC ≤ 15 + various filters	≈ 8000 compounds Reference : no Analogs : yes	Polar hydrogens
	Linking	1		Ki ≈ 60 μM LE=0.58	amide	Linking Mass 150 Da	Canonical amide "C(=O)[NH]" preserved	≈ 650 compounds Reference : yes Analogs : yes	Polar hydrogens
		3		LE=0.27	amine or amide or sulphonamide				
	Targeted Reference Lead	4		Ki ≈ 0.002 μM LE=0.49					
Hsp90	Linking	11		IC50 ≈ 1500 μM	amine	Linking Mass 250 Da	/	≈ 300 compounds Reference : no Analogs : yes	Polar hydrogens
		12		IC50 ≈ 1000 μM	ester				
	Targeted Reference Lead	13		IC50 = 1.5 μM					
AChBP	Growing	21		pKi = 5.3 LE=0.43	amine	Growing Mass 200 Da HCNOSP	HAC < 14 + charged amine + various filters	≈ 4500 compounds References : yes Analogs : yes	Several sidechains + polar hydrogens
	Targeted Reference Leads	22		pKi = 7.5 LE=0.41					
		23		pKi = 7.0 LE=0.37					
AlSynth	Growing	31		pKi = N/A	urea	Growing Mass 200 Da	/	≈ 1550 compounds References : yes Analogs : yes	/
	Targeted Reference Lead	32		pKi = 7					

[a] It includes a 2D depiction of fragments with their binding data. Information about used linker database, employed RECAP flags, generated and screened ligands, filters, and flexible binding site moieties are provided, too.

specify a linker database, the considered RECAP flag(s), above-mentioned structure file(s) for precursor(s), and linkage points.

Initially, a "chimera" ligand is built by plugging in the linker group into the reference molecular graph and submitting it to the

ChemAxon conformer plug-in in order to obtain one relevant conformer thereof. It however stores the initial fragment coordinates as data fields to be used by S4MPLE in order to override the new coordinates produced by the ChemAxon tool. The user may specify if the major microspecies based on the ChemAxon pK_a plug-in should be computed. Eventually, hydrogens are added, and atomic partial charges (Gasteiger[65]) and various properties (heavy atom count, mass, number of rings, etc.) are computed and saved as an annotated MDL .sd file. The last ligand preparation step consists in the assignment of GAFF atomic types using antechamber[66] and their check with parmchk from the same software suite (AMBERTools[67]).

Both linking and growing strategies may generate a huge number of putative compounds. Common filters (e.g., heavy atom count HAC, rotatable bond count, ring count, acceptor count, donor count, chiral center count—see cutoff values in Table 1) are used to reduce the size of the final library to screen and to discard non-drug-like molecules.

2.3. Virtual Screening Protocol. S4MPLE is subsequently used to simulate the compounds within the binding site, ignoring the DoF of precursor fragments. These have by default a "fixed atom" status, like the majority—or totality—of site atoms. It is a good strategy, however, to unlock the immediate neighborhood of the linkage point(s) in precursor fragments, as local geometric rearrangements in the bound fragments may occur upon linking. In growing (Figure 1, top), S4MPLE will break down the

unlocked part of the ligand into pieces on which genetic operators will apply. This is an unambiguous operation because for each rotatable bond the molecular moiety not including the fixed atoms will be taken. If a linker, however, is spanned between two fixed fragments (Figure 1, bottom), S4MPLE cannot detect explicit DoF to the linker part—either moiety around a linker bond contains fixed atoms and cannot qualify as mobile pieces associated to an explicit DoF. This issue is addressed similarly to the intra-cyclic conformer sampling problem: a bond at the interface fragment-linker is automatically selected and formally "broken". This allows S4MPLE to detect mobile pieces ranging from the other fixed fragment atoms to this "broken" bond. There will be no explicit DoF associated to the "broken" bond, but its end is implicitly subject to all the moves of the detected pieces. Because the bond is formally broken but not removed from the molecular energy expression, its harmonic terms (bond stretching and angle bending) will be minimized in linker poses respecting these covalent constraints. Because S4MPLE includes a chirality inversion penalty term, the "broken" bond may include chiral atoms as well.

Genetic Algorithm-driven sampling consisted, for all simulations with rigid binding sites, in 200 generations of 20 individuals with all other parameters set to their default values.

In order to probe for putative water-mediated interactions (Acetolactate Synthase study case), no change of protocol is needed: it suffices to append the desired number of explicit water molecules (two, in this case) at the end of the multimolecule .sdf file. Albeit in the previous work we had reported specific protocols aimed to enhance the sampling of ligand-water-site bridges,[1] these were not used here (by default, waters tend to be positioned such as to preferentially form hydrogen bonds with ligand atoms or other waters, the natural "hot spots". No explicit hot spot list was provided in any reported simulations. Please refer to the cited publication and the S4MPLE user guide for more information). However, in order to account for the increased problem space, the 200 generation simulation was run in duplicate. Also, a duplicate 200 generation run without waters was performed in parallel, the goal being to compare the net impact of including explicit waters on the pertinence of the virtual screening.

When binding site flexibility was enabled (Acetylcholine Binding Protein study case), the sampling effort was extended to 400 generations. Like previously outlined,[1,59] a pretty large binding site is defined from the reference X-ray structure using the same workflow.

The sampling stage is followed by an optimization of all kept poses (all non-redundant poses at given interaction fingerprint dissimilarity threshold *minfpdiff*, within a given energy window +30 kcal/mol), now unlocking all the ligand DoF. Growing runs employed a *minfpdiff* = 0.005, while linking used a smaller *minfpdiff* = 0.001 because both ends are restrained. The fine minimization consisted in a systematic optimization of hydrogens from hydroxyl groups of the considered system (ligand and binding site, when unlocked), followed by the "exhaustive energy minimization" operator.[59] Figure 2 describes the full virtual screening protocol.

Given the initial fixing of precursor atoms, the final ligand conformers may retain an important intra-molecular strain component—all subsequent optimization notwithstanding. Therefore, scoring will not consider only ligand-site interaction terms but also include an estimate of this intra-molecular strain. The energetic criterion considered here thus equals the potential energy of the complex minus the potential energy of the best
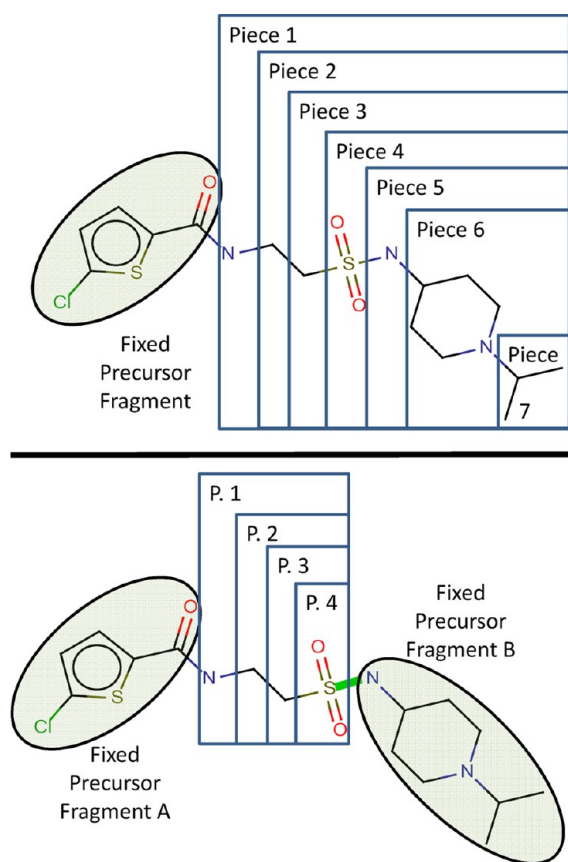

Figure 1. Automatic detection of moving pieces used in the evolutionary sampling routine. Top: Growing scenario (fixed precursor fragment). Pieces are hierarchically defined with respect to rotatable bonds, such as to include only moving atoms (here, amide bonds are considered rotatable as well). Bottom: Linking scenario, with the formally (green) "broken" bond. Moving pieces are all anchored to fragment A and end at the "broken" bond.
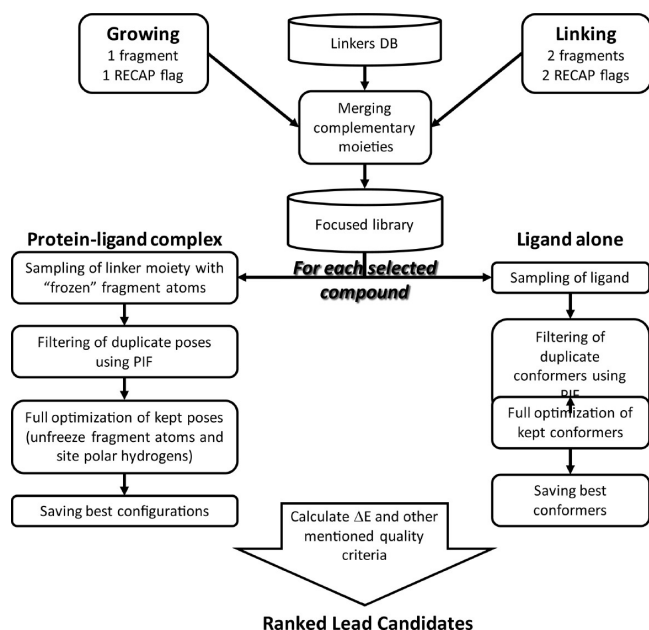
Figure 2. Workflow of the virtual screening protocol used in all studies.

conformer of the free ligand (the latter being obtained by a rapid run, consisting in 100 generations, in absence of the protein site). This would also include intra-site strain energies if site DoFs are enabled. With explicitly added water molecules, the basis energy level of waters in the free state is constant, so needs not be subtracted from the complex energy in as far as this score only serves for ranking/prioritization of candidates.

The final prioritization of the candidates is based on multiple criteria, the main being the force-field energy difference defined above. In the ideal case, this FF-based energy should discard bad linkers (constrained configurations, systematic clashes, etc.). Furthermore, ligands in which the initial fragments quit their initial poses, as soon as their DoF are unlocked at the final minimization stage, are unlikely to represent potent binders—or, at least they are outside the applicability domain of FBDD. A RMSD criterion is employed in addition to the energy: molecules with RMSD shifts over initially fixed precursor atoms beyond a user-defined threshold (respectively 0.5 and 1.0 Å in growing and linking procedures) are automatically discarded. Eventually, poses with obvious local structural flaws (that are but slightly penalized in terms of energy) are also discarded: here, this concerns molecules for which the best conformation within the binding site contains unlikely amide configurations (deviating by >10° from trans). Combining all these criteria is expected to rank interesting linkers in the top hit-list.

2.4. Data Sets. Several lead optimization strategies (starting from a fragment-like ligand) or explicit growing/linking approaches are extracted from literature, the goal being to virtually reproduce these success stories. Various targets emerged from this selection: activated Factor X (FXa), Heat Shock Protein 90 (Hsp90), and Acetylcholine Binding Protein (AChBP). Acetolactate Synthase (AlSynth) was used to exemplify growing runs with explicit waters. Table 1 sums up, for each data set, the main information about inputs, filters, and used options.

2.4.1. Factor Xa (FXa). This serine protease, involved in the coagulation cascade, has been intensively studied for decades, and several tens of complexes are deposited in the Protein Data Bank[68] (PDB). FXa is the activated form of the enzyme, and its inhibitors are putative anticoagulants drugs. A large variety of

molecular scaffolds are present in these active compounds. Nazare et al. recently exemplified the affinity gain due to successful linking[14] from the deconstruction of small FXa inhibitors of high affinity.[69] The reference drug-like compound **4**, directly obtained from the merging of two low affinity fragments, has an impressive LE of 0.49.

The present is a retrospective study, where both in silico growing and linking strategies are applied to "precursors" from the fragmented reference ligand. Canonical RECAP flags amine, amide, and sulphonamide being present, the herein envisaged virtual evolution strategy to recreate the potent inhibitor **4** makes perfect chemical sense. The starting materials are one X-ray structure (4A7I) and binding data. Two growing and one linking protocols are carried out:

- Growing I (small): optimization from fragment **2** (LE = 0.36) with the sulphonamide extension. The chlorothio-phene-amide moiety is fixed; other atoms of fragment **2** are mobile.
- Growing II (large): evolution from fragment **1** (LE = 0.58) with the amide flag. Chlorothiophene is the fixed moiety of this fragment.
- Linking: connect fragment **1** (flag amide), a comple-mentary spacer and fragment **3** (flags amine, amide, or sulphonamide). Both original rings (chlorothiophene and isopropyl-piperidine) are kept fixed at the first stage.

2.4.2. Heat Shock Protein 90 (Hsp90). The Heat Shock Protein 90 (Hsp90) is a molecular chaperone involved in the refolding of several client proteins related to cancer.[70,71] Hsp90 is overexpressed into some cancerous cells and makes them more resistant to treatments, so its inhibition should reduce their life expectancy. The popular Hsp90 target has been intensively studied within FBDD projects,[72−76] and numerous complexes have been deposited in the PDB. Among them, several structures contain two fragments (concomitant binders) in the same binding site, enabling a linking strategy. The X-ray structure 3HZ5 contains a larger ligand (**13**), which results from the linking of N-dimethyl-adenine (**11**) and a pyrazole-derivative (**12**) fragment, after removal of the ester group of the latter (because not involved in favorable site contacts). These two non-competitive fragment-hits are present in the structure 3HZ1 and are linkable because the reactive centers are not sterically hindered. This technique was employed by Barker et al.[73] to combine the fragments **11** and **12** into ligand **13**. Each individual fragment exhibits very low potency (IC$_{50}$ in low mM range), whereas the linked inhibitor gained 3 orders of magnitude (IC$_{50}$ = 1.5 μM).

A virtual linking simulation is performed for the same purpose. The binding site is taken from the 3HZ1 structure. The set of anchor atoms to keep fix at the first sampling stage comprises the purine and pyrazole-derivative rings.

2.4.3. Acetylcholine Binding Protein (AChBP). The Acetyl-choline Binding Protein is released into the synaptic cleft in order to capture the acetylcholine neurotransmitter. The direct consequence is the modulation of the synaptic transmission. In practice, this protein serves as model to study related membrane proteins such as ligand-gated ion channels.[77] Using AChBP as the target, Edink et al. described a fragment optimization (growing) resulting in conformational changes in the binding site.[78] As in the Hsp90 study, the starting materials are binding values and structural data for both fragment and evolved ligands. The 2Y54 structure includes the fragment **21** that corresponds to the Murcko scaffold[79] of cocaine. The visual analysis of the complex shows that the small

organic compound binds to the bottom of cavity, inside an aromatic cage perfectly suitable for binding hydrophobic cations such as choline or tropine. The comparison of different X-ray structures, including either fragment **21** (pK_i = 5.3) or a reference potent inhibitor (α-lobeline, pK_i = 8.6), suggested one optimization way, while highlighting that conformational changes must be mandatory to accommodate ligands evolved from fragment **21**. As expected, the designed molecules bind to AChBP and the X-ray structures of these complexes (2Y56/2Y57) showed one main modification in the binding site conformation: another Y91 rotamer giving access to a new subpocket, occupied by the added chemical groups. Compounds **22** and **23**, optimized from fragment **21**, have respective pK_i values of 7.5 and 7. In addition to a thermodynamic analysis of binding profiles, the authors demonstrated that fragment evolution can (obviously) impact the pocket conformation.

From an in silico point of view, a retrospective growing analysis based on these data appears very interesting and challenging at the same time because flexibility of the binding site must be enabled for relevant reasons (experimentally confirmed binding site rearrangements). Although there is only one large shift between the 2Y54 and 2Y56/2Y57 structures (Y91 rotamer), all side chains around the ligand are unlocked (Y53, Y91, W145, Y186, and Y193) in order to make more challenging and reliable simulations. In this growing case, based on the amine RECAP flag, no additional unlocked atoms are needed because the evolution process directly starts from the bridged ring of fragment **21**. The binding site is taken from the 2Y54 X-ray structure. The resulting database of grown analogs is huge; thus different filters based on molecular properties are used to create a more tractable set, as in the large FXa growing II protocol. Because the key role of the positive charge carried by the ammonium group is well acknowledged, analogs in which the starting nitrogen loses its basic character by coupling to the linkers are discarded before docking/sampling in order to save time. This filtering is straightforward based on the predicted protonation state in the major microspecies returned by the ChemAxon pK_a plugin.

2.4.4. Acetolactate Synthase (AlSynth). This study case has been chosen because the targeted reference ligand **32** shows two key water-mediated interactions with the site (PDB 1T9B[80]), all while allowing a straightforward retrosynthetic cut of the species (at the level of the urea group). There is no experimental knowledge of the binding status and position of the starting fragment **31** (4-methoxy-6- methyl-1,3,5-triazin-2-yl)-urea; its location seen in the final lead is therefore assumed.

The carbonyl-sulphonamide group is expected to have a pK_a value of about 4 (according to the ChemAxon pK_a prediction tool[81]), and the ligand is indeed considered under its ionized anionic form in docking benchmarks (Astex data set[82]). However, the position of the first key crystallographic water— perfectly symmetrically located at the convergence point of a bidentate hydrogen bond with both the N−H groups of the urea—suggests that within the active site the effect of the protein has modified the local pK_a value as to prefer the neutral form. It was known from our previous results[1] that S4MPLE was able to dock the entire ligand (at that time taken as the anionic species $C(=O)[N^-]SO_2$ together with its two crystallographic waters, leading to a predicted geometry within the top 10 best ranked ones in very close agreement with the PDB structure. However, the placement of the urea-binding water was no longer in a symmetric position between the urea nitrogens. For simplicity, the neutral form of the ligands was imposed in these simulations.

The first above-mentioned crystallographic water bridge involves atoms of the prepositioned root fragment and is expected to occupy that position regardless of the nature of the growing moiety. It could have been fixed and integrated to the site, but was nevertheless allowed to move freely, serving as an implicit control of sampling quality.

The second water mediates the interaction between the sulphonyl oxygen and the cation of K251. This is the water molecule that may eventually participate (if possible, as in the case of the chlorophenylsulphonyl moiety) in non-covalent "growing", completing the connected linker and granting an extra stabilization for the binding. Otherwise, it is allowed to adopt another favorably interacting position with the site and/or the ligand.

The goal here is to retrospectively find the chlorophenylsulphonyl moiety, out of a large collection of alternatives, and to show that its finding is significantly enhanced when accounting for explicit waters.

3. RESULTS AND DISCUSSION

All results relative to the different growing and linking protocols are reported and discussed below. These should answer the previously issued questions (see the end of section 1.4). The initial fixing of precursor fragment atoms is useful for several reasons. First, the goal consists in the optimization of an original hit already present in the binding pocket (experimental location, or obtained by docking). Therefore, only binding modes sharing the initial interactions are really interesting. Besides, in the context of linking, a suitable spacer must freeze original fragments into their relative positions when independently bound to the site. Finally, these constraints have the advantage to speed up the process compared to a full blind docking.

In prospective projects, it may be interesting to redock all selected compounds without any particular bias. If unbiased docking leads to a same optimal pose as the growing/linking strategy, this may be interpreted as an additional argument augmenting the confidence in the success of that particular compound. However, if unbiased docking reveals lower energy poses with original fragments in unexpected locations, this may signal a force field artifact (fake minimum of the energy landscape, not accessible under growing/linking constraints). Alternatively, it may simply reveal a better overall pose with fragments choosing only slightly suboptimal alternative binding pockets. However, practice shows that observation of known key interactions in a ligand-site complex is often a more reliable success indicator than favorable scoring function values—this is why interaction fingerprints were first introduced.[83,84] Because in the herein addressed retrospective studies the experimental poses of final ligands are known, no final unbiased docking was undertaken to illustrate the expected direct anchoring of the ligand its site. Having the native or native-like linkers or growth fragments highlighted as preferred choices out of the large set of decoy substructures counts here as success. The question whether these best options are good enough to yield a potent ligand is already answered.

3.1. Factor Xa. This target is widely studied here in three protocols (two growing and one linking). Different fragments are used as inputs, but the ultimate goal is always the same: generating the high LE reference **4** and/or close analogs and recognizing them as putative "top hits" among all the other generated growth/linking products. Departure fragments **1** and **2** have low but measurable K_i, therefore they should be discovered during a typical fragment screening campaign. Growing protocol II

G

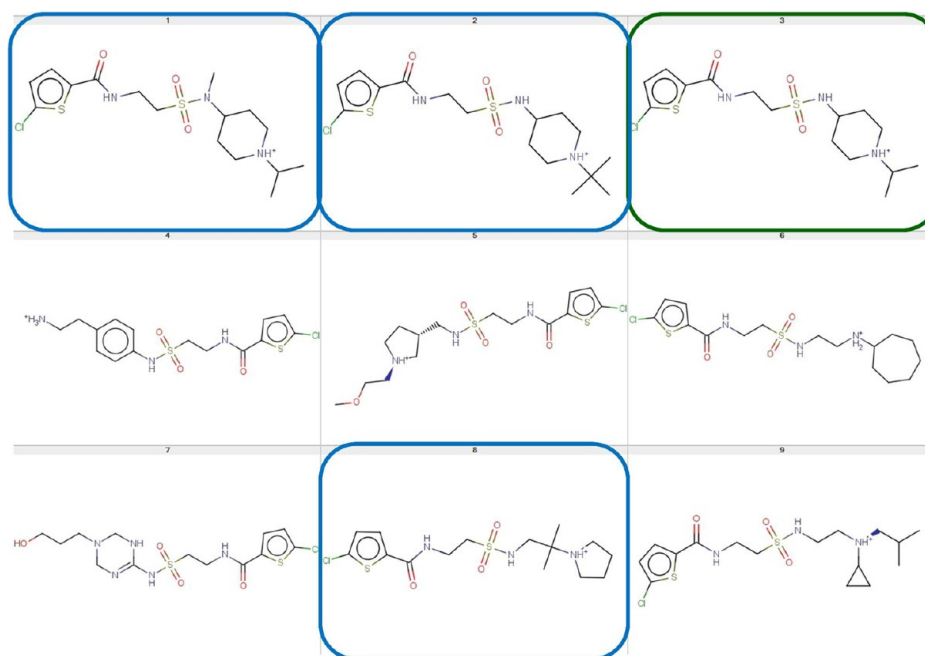dx.doi.org/10.1021/ci4000163 | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

Figure 3. Top nine molecules from the FXa growing I virtual screening. Reference compound and closest analogs are respectively highlighted in green and blue.
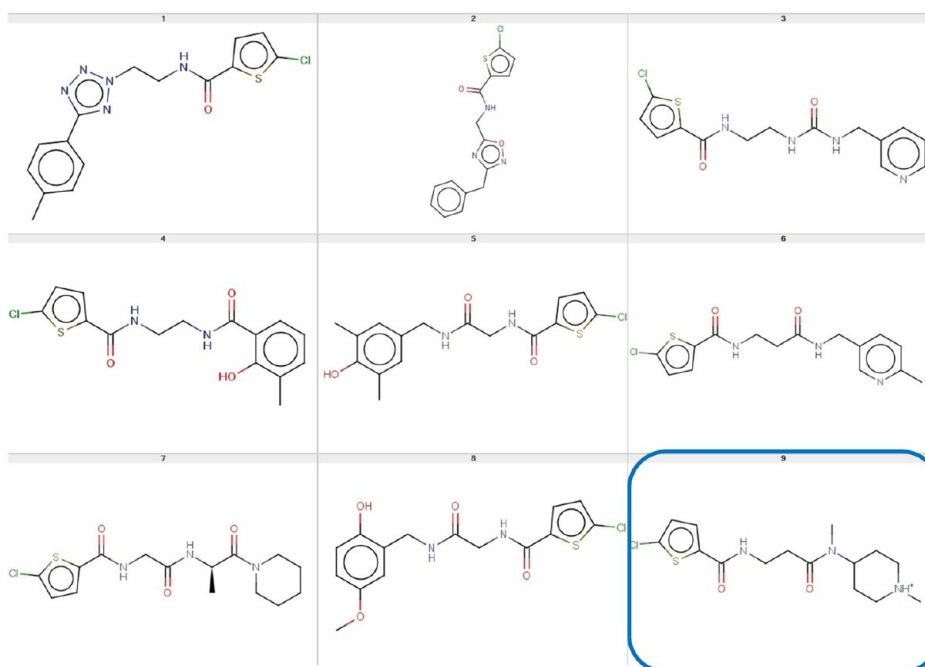

Figure 4. Top nine molecules from the FXa growing II virtual screening. Closest analog is highlighted in blue.

mimics a real situation, where only highest LE fragments are selected for the subsequent optimization stages. While the protocol II has stronger arguments, it is really more challenging because a larger moiety must be designed via the growing strategy. Conversely, strategy I is merely required to add an isopropyl-piperidine group to the original fragment. The virtual linking approach, realized here for the POC, would probably not have been undertaken in real conditions because fragment **3** has a very low affinity/efficiency (LE = 0.27). Unfortunately, the binding modes of fragments **1**, **2**, and **3** alone were not experimentally solved, contrary to that of reference product **4**. This information has been confirmed by the first author of the

study of interest.[69] However, various known inhibitors contain the chlorothiophene substructure (**1**), and the latter always binds the S1 pocket as for ligand **4**. Besides, docking (with S4MPLE) of fragments **1** and **3** always led to the supposed binding mode each time with a large energy difference with respect to the second pose. Thus, the assumption that individual fragments have the same binding mode as in the grown/linked potent inhibitor appears as a reasonable hypothesis for these retrospective in silico studies, although several examples from FBDD projects highlighted the opposite case.[73]

3.1.1. Growing Protocol I (Small). This growing protocol, simpler than the others, is the first step of the POC workflow:

is the strategy able to prioritize the chemical group seen in molecule **4** matching the S4 pocket of FXa? The fact of unlocking some significant DoF in the precursor fragment **2** (the "sulphonamide spacer") opens the theoretical possibility to reach various alternative pockets with the grown chain (a correctly prepositioned rigid sulphonamide would have obviously oriented the substituents toward S4 making this challenge too easy). Over the 4000 grown ligands, the top nine compounds, selected after pruning those with unacceptable RMSD shifts and distorted amide conformations, contain the known inhibitor (compound **4** ranked as #3) and several very close analogs (ranked as #1, #2, and #8). These top nine molecules are depicted in the Figure 3. Although only moderate-sized chemical groups were sought here (cyclic aliphatic ammonium), the best binding mode perfectly matches the experimental one (RMSD < 0.5 Å). On the basis of these encouraging results, this first growing protocol is validated.
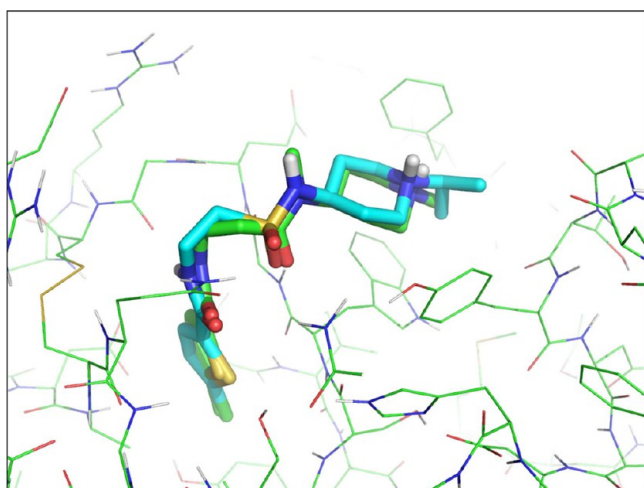

Figure 5. Superimposition of compound #9 (methyl-amide) to the high LE inhibitor **4** (sulphonamide) from the FXa growing II virtual screening.

3.1.2. Growing Protocol II (Large). In this case study, the preliminary in silico growing stage is expected to connect the ethyl-sulphonamide-piperidine-isopropyl group (15 heavy atoms) to the starting fragment **1** to exactly reproduce the reference **4**. Unfortunately, this inhibitor is not automatically generated. An analysis of the large growing database revealed that the desired group ($_{amide-flag}$-ethyl-sulphonamide-piperidine-iso-propyl) was simply not among the options retained when fragmenting the ZINC-clean collection. This is a clear limit of the knowledge-based approach used here: although containing more than 9 million compounds, the ZINC database only represents a fraction of the huge drug-like chemical space. Despite the absence of desired ligand (**4**), the virtual screening was undertaken nevertheless because potentially interesting compounds including an ethyl-sulphonamide spacer or a terminal hydrophobic cation were generated. Because contributions of non-specific site-ligand interactions to the binding energy scale up with putative ligand size, this growing was confined to fragments within a size window of 12 to 15 heavy atoms. Within these specifications, about 8000 molecules were finally screened, and several putative interesting compounds emerged among the top 100 hits. As before, the top nine compounds are reported (Figure 4). In particular, the ninth hit appears pretty close to the active inhibitor **4** with respect to two criteria:

- Chemical structure: Amide linkage substitutes the sulphonamide spacer and a terminal N-methyl replaces the N-isopropyl one.
- Binding mode: Superimposition of experimental binding mode of the reference and of predicted hit 9 is depicted in Figure 5.

Figure 6 shows some more hits within the top 100, all of which may be considered rather similar to **4** because they all include either an ethyl-sulphonamide spacer followed by a terminal hydrophobic group or a terminal charged hydrophobic structure like the isopropyl-piperidine.

Finally, from an energy point of view, the top ranked ligand exhibits a higher FF energy difference (less favorable) than the
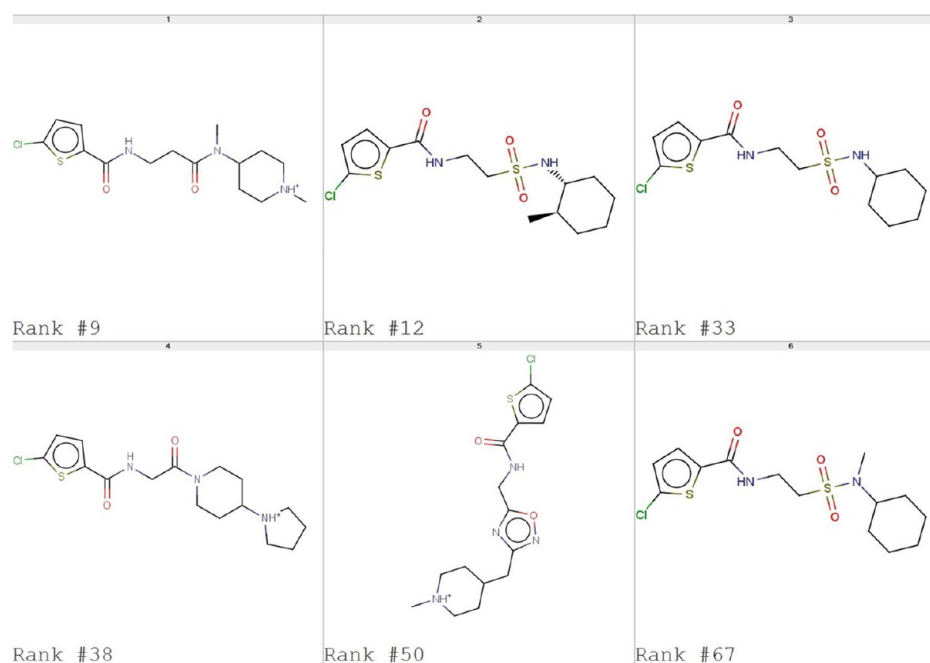

Figure 6. Most similar compounds to reference **4** within the top 100 hits from the FXa growing II virtual screening.
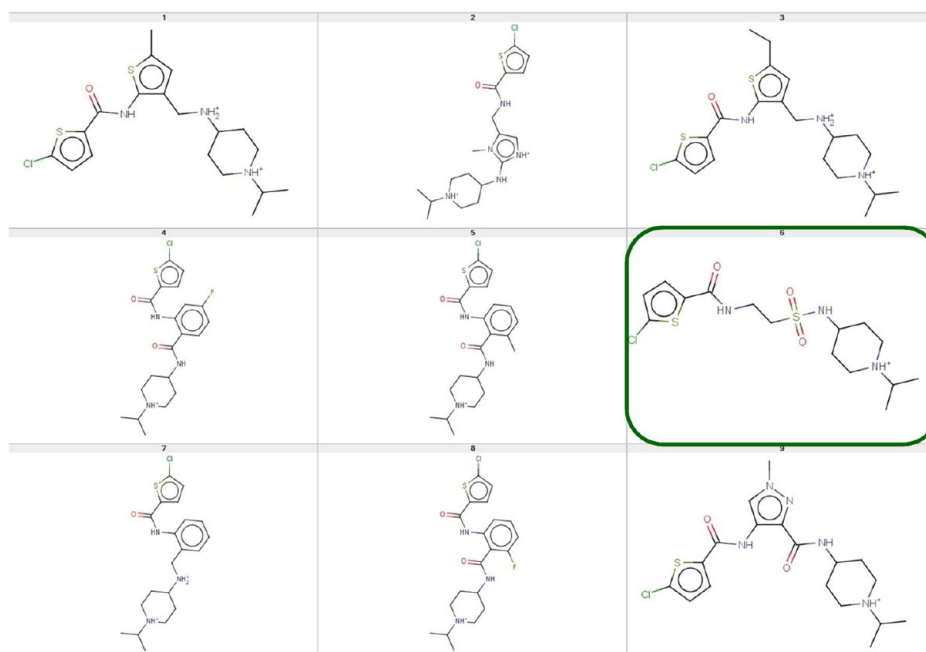
Figure 7. Top nine molecules from the FXa linking virtual screening. Reference compound is highlighted in green.
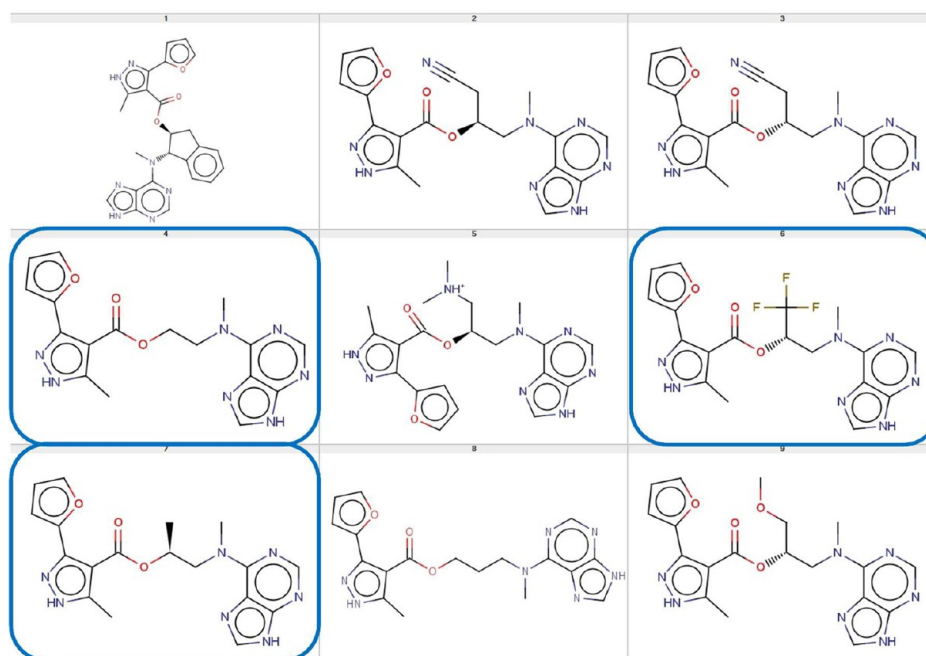

Figure 8. Top nine molecules from the Hsp90 linking virtual screening. Closest analogs are highlighted in blue.

reference compound **4** simulated in the same conditions. If the needed moiety would have been present at growth stage, **4** would have been top ranked with respect to all herein retained candidates. Although not perfect, the results of this larger virtual screening highlight valuable perspectives for prospective growing studies.

3.1.3. Linking Strategy. This first linking approach, consisting in the enumeration and assessment of about 650 compounds, generated and retrieved the expected ligand in the top nine ligands (reference compound **4** ranked as #6). No specific filters were employed here to reduce the set of compounds to screen, except for the preservation of a canonical amide C(=O)[NH] (with one polar H, at least). The sulphonamide spacer occupies

its expected location. However, comments about the accuracy of retrieved binding mode are less relevant in this situation because both ends are constrained in the linker sampling stage (before the relaxation of the full ligand). Compared to the growing protocol I, only few analogs are retrieved here—expectedly, because only the chemical space of the linker moiety is subject to exploration. One close analog, with an additional carbon in the linker, is ranked #90. Visually, this new carbon displaces the sulfonamide group that loses its favorable interactions (hydrogen bonds) with the site, hence a logically worse ranking. Figure 7 displays the top nine results of this linking simulation. Surprisingly, there are several molecules including two ammonium ions according to the ChemAxon major microspecies plug-in. This may be correct

for the unbound ligands but may change due to pK_a shifts upon binding. However, such subtle effects are beyond the reach of FF-based modeling. Also, the relative gain in binding energy upon presence of a second cation may be an artifact due to ignoring counterion effects. All in all, chemical common sense recommends caution in accepting these species as putative actives.

To conclude, the linking procedure automatically generated and top ranked the desired inhibitor **4** from fragments **1** and **3**, thus representing a success.

3.2. Heat Shock Protein 90. Conversely to the FXa studies, where hypothesis are made about the fragments binding modes, X-ray data here give access to simultaneous binding modes of non-competitive fragments (3HZ1) in addition to the evolved compound (3HZ5). However, the final compound does not fully include the starting fragments: the masked acidic group present in the pyrazole-derivative **12** disappears from the optimized ligand. Authors highlighted the fact that the ester group does not

make tight interactions with the site. Besides, this chemical function is known to be frequently cleaved in vivo. Therefore, this group was removed in the linking process and was replaced by a simple alkyl spacer of length suggested by molecular modeling. The direct consequence is that our procedure can only generate more or less close analogs including the ester function but not the exact known reference **13**. Given that only few data (including X-ray support for individual fragments and evolved compound) relative to linking optimization are available, this retrospective linking study has been undertaken anyway.

Ester-containing analogs are retrieved at ranks #4, #6, and #7 (Figure 8). Ligand #4 looks like the closest analog of the reference because its spacer is very simple (not ramified) and does not contain any stereocenter, conversely to linkers of #6 and #7. Empirically, a good superimposition between the active (**13**) and these suggested molecules can be observed, preserving the hydrogen bonds network between the purine ring and the binding site (Figure 9). Interestingly, the protocol favored molecules (eight among the top nine) with a topological distance of 6 between their heterocycles (pyrazole and purine), as in the known inhibitor **13**.

Despite the inability to generate the reference inhibitor **13**, which does not exactly incorporate original fragment hits, this in silico linking procedure generated compounds with rather similar linkers (same length as in the active ligand, but including an ester function) within the top hit list.

3.3. Acetylcholine Binding Protein. A strong feature in S4MPLE, exemplified by this case, is that growth/linking may be performed all while considering binding site flexibility (during the sampling and/or the final refinement), whereas many state-of-the-art tools still operate today on rigid sites only. Most issues mentioned in the introduction are addressed by this simulation: design of known inhibitors, sampling of generated molecules (e.g., picking up the expected interactions between the site and known ligands), ranking of references in the top hit list, and prediction of binding site conformational changes. The optimization of fragment **21** into actives **22** or **23** consists in adding a pretty simple group (e.g., phenyl with two spacer
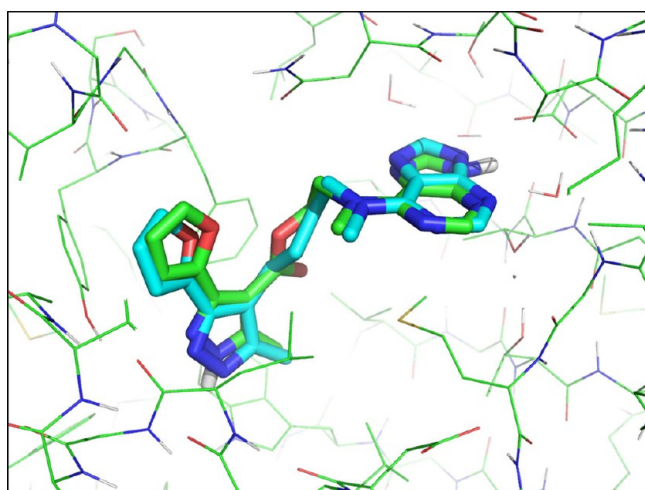

Figure 9. Superimposition of the reference inhibitor **13** (green) to its closest screened analog (ranked as #4 and displayed in blue).
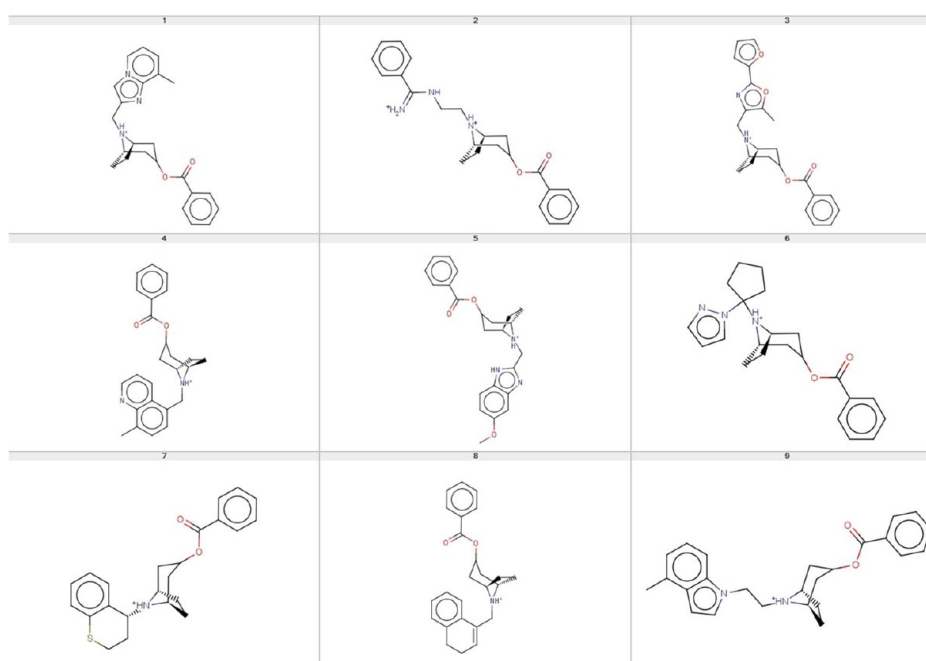

Figure 10. Top nine molecules from the AChBP growing virtual screening.

carbons for molecule **22**). As expected, these molecules are automatically generated from the fragment **21** using the amine RECAP flag. The top nine retrieved molecules are displayed in the Figure 10. By contrast to previous successful cases, molecules **22** and **23** are not ranked in the top hit list (e.g., top 1%) in this more challenging study, involving about 4500 screened entities. However, ligand **23** is ranked among the top 4%, whereas ligand **22** appears in top 10%, and one close analog of **22** (tolyl group substitutes the phenyl ring) is ranked in the top 3%. Figure 11
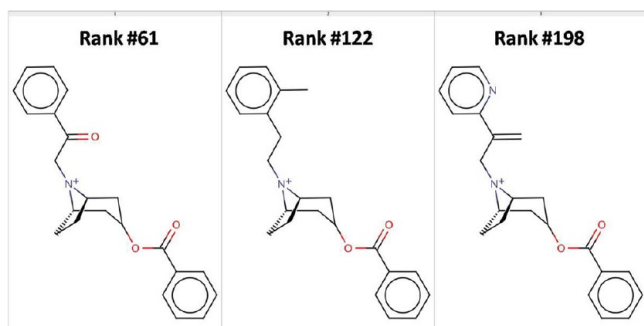

Figure 11. Most similar compounds to references **22** or **23** within the top 5% hits from the AChBP growing virtual screening.

shows some hits within the top 5%, all rather similar to references **22** or **23**. Finally, the expected new rotamer of the tyrosine (Y91) is successfully predicted for sampled reference molecules. The additional hydrogen-bond between compound **23** and binding site backbone is reproduced, while the other four flexible side chains correctly fall back into their original conformations. Figure 12 illustrates starting materials (experimental binding modes of fragment **21** and ligand **23** within their respective binding sites) and sampled geometries of reference **23** into the partly flexible 2Y54 binding site.

Although the ranking appears a bit less efficient in that specific case (pessimistically assuming that all compounds ranked above the expected references are false positives), most previously introduced issues are favorably addressed by this in silico growing protocol with partial binding site flexibility.

3.4. Acetolactate Synthase. If this simulation is run as a classical water-free growing process, like in examples above, the targeted reference ligand **32** would have been ranked as the 62nd most promising moiety to connect to the departure triazinyl-urea fragment **31**, out of about 1550 candidates (i.e., within top 4%). With explicit waters, however, its ranking is boosted to position 6 (i.e., within top 0.4%). Therefore, inclusion of explicit-free waters appears as an important factor in order to highlight the suitability of the chlorophenylsulphonyl moiety as a correct choice for growing. Furthermore, the waters are both found at expected locations, all while the geometry of the bound ligand is perfectly reproduced (Figure 13). The best nine ligands selected by the explicit-water simulation are shown in Figure 14. Intriguingly, the overall best ranked compound, having N-bound phenylalanine as selected growing moiety, is common to both explicit-water and water-free runs. The negative carboxylate therein forms a strong salt bridge with the lysine (albeit rather solvent exposed), and waters are both located in the neighborhood of the carboxylate. It is an illustration of a possible compound in which the water-mediated interaction would be replaced by a direct contact, showing that if such alternatives were present among the possibilities opened by the growing moiety database they could be found as well.
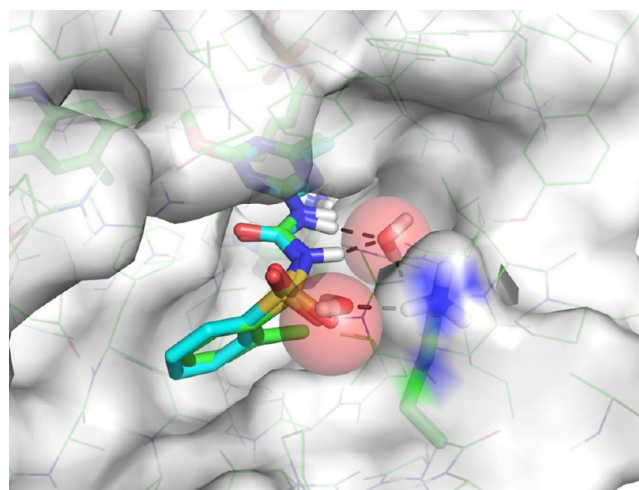

Figure 13. Superimposition of experimental and predicted geometries for the reference compound **32** from the AlSynth growing virtual screening. Predicted waters locations are shown as sticks, and experimental ones are displayed as spherical.

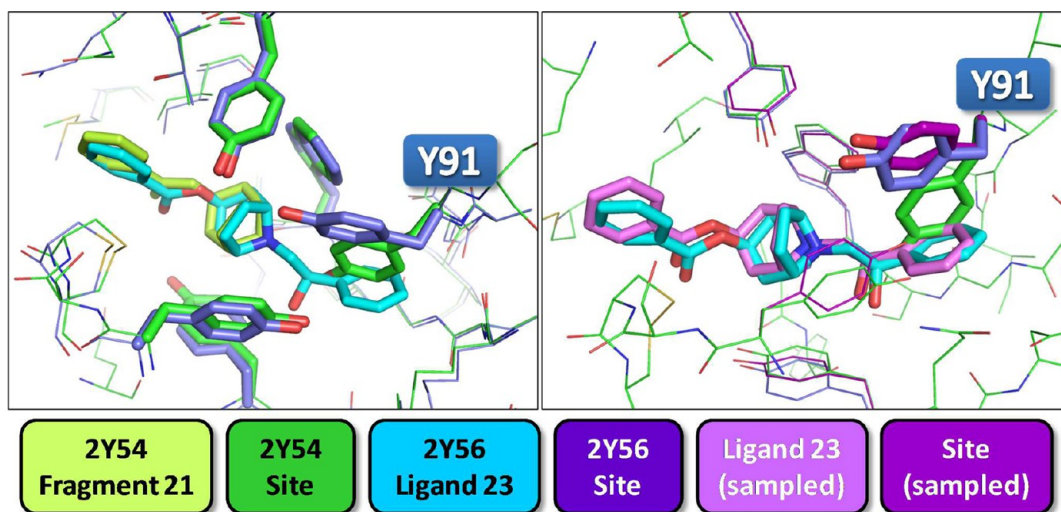| 2Y54 Fragment 21 | 2Y54 Site | 2Y56 Ligand 23 | 2Y56 Site | Ligand 23 (sampled) | Site (sampled) |

Figure 12. Superimposition of experimental structure (2Y54 vs 2Y56, left) and resulting sampled configurations (right). Albeit the initial site geometry was taken from 2Y54, the sampled geometry matches closely the expected 2Y56. Conformational changes of Y91 are highlighted.
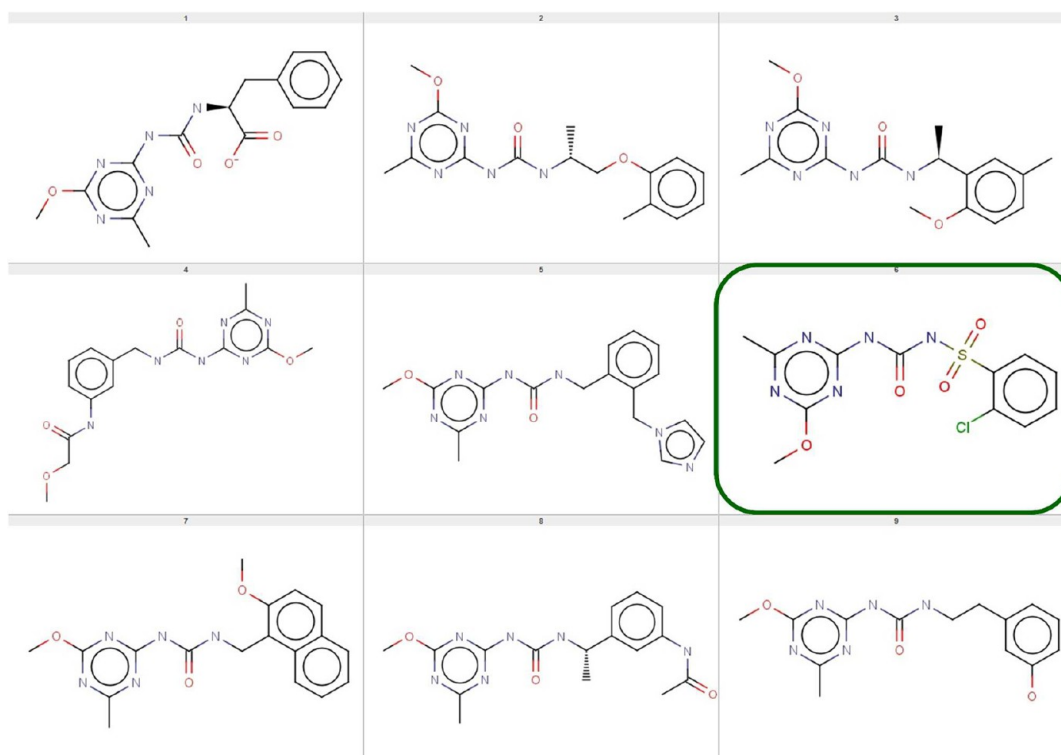
Figure 14. Top nine molecules from the AlSynth growing virtual screening. Reference compound is highlighted in green.

Unfortunately, we have no knowledge of the activity level of this interesting derivative. Finally, the top 50 selected ligands in both runs have nevertheless significant overlap: common species are molecules with hydrophobic groups (mainly interacting with V191, P192, A195, A200, F201, and K251), obviously well ranked irrespective of the use of explicit waters.

4. CONCLUSION AND PERSPECTIVES

On the basis of several retrospective studies, this in silico fragment-to-lead protocol involving JMolEvolve and S4MPLE is validated by its ability to automatically generate expected ligands (and/or close analogs) from starting fragments, to rank them with some effectiveness among a set of decoy compounds, to reproduce their experimental binding mode, and to predict slight site conformational changes when binding site flexibility enabled. Furthermore, S4MPLE—initially designed in view of conformational sampling and docking—is not only well able to take charge of problems typically addressed by specific FBDD/DND computer programs but allows going beyond classical applications. It may easily include site flexibility, but foremost, it is able to consider non-covalent alternatives to classical growing (i.e., including non-covalent partners such as water molecules to mediate favorable contacts with the site). This allows priority retrieval of growing moieties that would not be top-ranked under default conditions because their complementarity to the site is less apparent if bridging water molecules are not present. All this is achieved without any prerequisite knowledge of the expected water positions and without any specific alterations of the protocol: the only user-defined parameter is the number of explicit water molecules to consider (where adding too many should not impact on the found binding modes, albeit it would impact on the convergence rate).

This optimization strategy is currently being applied in prospective FBDD projects (work in progress at NovAliX, Strasbourg). Even the herein used minimalistic but robust linking/growing strategy may

certainly produce novel compounds if provided with original starting fragments and linkers. Assessment of synthetic accessibility was not the subject of this structure-based study, although the RECAP-based procedure did not generate odd structures. Nevertheless, future work will focus on that important point in order to increase the usefulness of our in silico FBDD protocol.

S4MPLE-driven docking will at least ensure that retained linking/growing options match well the shape of the active site and are rather void of intra-molecular strain. Will they be active? This is the key question. Depending on the accuracy of the herein used AMBER/GAFF force field, of the additional desolvation and contact terms, and on entropic aspects no one really knows how to accurately simulate, etc. This paper evidenced the ability of the generic tool S4MPLE to address atypical problems in the sense that "docking" of linkers is driven by both covalent and non-covalent constraints, all while supporting side chain flexibility. The versatility of S4MPLE allows it to easily adapt to the "philosophy" of FBDD: specifically search for docked solution in which reference fragments remain in the positions they would spontaneously adopt when unconnected. Most important is the ability of S4MPLE to prioritize groups that benefit from water-mediated interactions with the site. This formally exemplifies (as far as we can tell) an original growth strategy of a (ligand + mediating water) non-covalent complex by contrast to classical growth through covalent bonds only. In most cases, S4MPLE managed to highlight (i.e., to prioritize in the energy-ranked list) known binders out of many decoys and to discover native-like poses (including the last example with freely roaming water molecules). The good news—from an algorithm development point of view—is that the tool did not get lost in phase space and did not get blocked in local high-energy minima (no failure to "drag" growing/linking moieties of initially random coordinates into a constrained active site) but actually found the relevant minima. That, further-more, the minima fortunately coincided with experimental input is

mainly the merit of AMBER/GAFF, partially enhanced by specific contributions.

Because of the full control over considered DoF, S4MPLE is a versatile tool covering from massively parallel computing of large ligands and small proteins (work in progress) to the medium throughput construction of reasonable growth/linker entities (at less than a CPU hour per ligand under a preliminary protocol employing a fixed number of evolutionary generations chosen such as to suffice for the most flexible among the encountered candidate compounds). Optimizing the termination conditions (rendering them dependent on the effective number of DoF and/or replacing the total generation number criterion by a maximal number of "stagnant" generations having produced no better offspring) must still be undertaken.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

S4MPLE (x86_64) version can be uploaded from http://infochim.u-strasbg.fr (see Downloads). Both compound files used in this work and the ChemAxon-driven growth/linking tools are available upon request (ChemAxon licenses needed by the end users). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: dhorvath@unistra.fr.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Hoffer, L.; Horvath, D. S4MPLE - Sampler for multiple protein−ligand entities: Simultaneous docking of several entities. *J. Chem. Inf. Model.* **2012**, *53* (1), 88−102.

(2) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent developments in fragment-based drug discovery. *J. Med. Chem.* **2008**, *51* (13), 3661−3680.

(3) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A rule of three for fragment-based lead discovery? *Drug Discovery Today* **2003**, *8* (19), 876−877.

(4) Hoffer, L.; Renaud, J. P.; Horvath, D. Fragment-based drug design: Computational and experimental state of the art. *Comb. Chem. High Throughput Screening* **2011**, *14* (6), 500−520.

(5) Neumann, T.; Junker, H.; Schmidt, K.; Sekul, R. SPR-based fragment screening: advantages and applications. *Curr. Top. Med. Chem.* **2007**, *7* (16), 1630−1642.

(6) Perspicace, S.; Banner, D.; Benz, J.; Müller, F.; Schlatter, D.; Huber, W. Fragment-based screening using surface plasmon resonance technology. *J. Biomol. Screening* **2009**, *14* (4), 337−349.

(7) Vivat Hannah, V.; Atmanene, C.; Zeyer, D.; Van Dorsselaer, A.; Sanglier-Cianférani, S. Native MS: An 'ESI' way to support structure- and fragment-based drug discovery. *Future Med Chem* **2010**, *2* (1), 35−50.

(8) Orita, M.; Warizaya, M.; Amano, Y.; Ohno, K.; Niimi, T. Advances in fragment-based drug discovery platforms. *Exp. Opin. Drug Discovery* **2009**, *4* (11), 1125−1144.

(9) Murray, C. W.; Blundell, T. L. Structural biology in fragment-based drug design. *Curr. Opin. Struct. Biol.* **2010**, *20* (4), 497−507.

(10) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **2004**, *9* (10), 430−431.

(11) Hann, M.; Leach, A.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3), 856−864.

(12) Murray, C. W.; Rees, D. C. The rise of fragment-based drug discovery. *Nature Chem.* **2009**, *1* (3), 187−192.

(13) Erlanson, D. A. Fragment-based lead discovery: A chemical update. *Curr. Opin. Biotechnol.* **2006**, *17* (6), 643−652.

(14) Borsi, V.; Calderone, V.; Fragai, M.; Luchinat, C.; Sarti, N. Entropic contribution to the linking coefficient in fragment based drug design: A case study. *J. Med. Chem.* **2010**, *53* (10), 4285−4289.

(15) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274* (5292), 1531−1534.

(16) Hajduk, P. J.; Sheppard, G.; Nettesheim, D. G.; Olejniczak, E. T.; Shuker, S. B.; Meadows, R. P.; Steinman, D. H.; Carrera, G. M.; Marcotte, P. A.; Severin, J.; Walter, K.; Smith, H.; Gubbins, E.; Simmer, R.; Holzman, T. F.; Morgan, D. W.; Davidsen, S. K.; Summers, J. B.; Fesik, S. W. Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.* **1997**, *119* (25), 5818−5827.

(17) Szczepankiewicz, B.; Liu, G.; Hajduk, P.; Abad-Zapatero, C.; Pei, Z.; Xin, Z.; Lubben, T.; Trevillyan, J.; Stashko, M.; Ballaron, S.; Liang, H.; Huang, F.; Hutchins, C.; Fesik, S.; Jirousek, M. Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *J. Am. Chem. Soc.* **2003**, *125* (14), 4087−4096.

(18) Law, R.; Barker, O.; Barker, J. J.; Hesterkamp, T.; Godemann, R.; Andersen, O.; Fryatt, T.; Courtney, S.; Hallett, D.; Whittaker, M. The multiple roles of computational chemistry in fragment-based drug design. *J. Comput.-Aided Mol. Des.* **2009**, *23* (8), 459−473.

(19) Hubbard, R. E.; Chen, I.; Davis, B. Informatics and modeling challenges in fragment-based drug discovery. *Curr. Opin. Drug Discovery Dev.* **2007**, *10* (3), 289−297.

(20) Hann, M.; Green, R. Chemoinformatics - A new name for an old problem? *Curr. Opin. Chem. Biol.* **1999**, *3* (4), 379−383.

(21) Schuffenhauer, A.; Ruedisser, S.; Marzinzik, A.; Jahnke, W.; Blommers, M.; Selzer, P.; Jacoby, E. Library design for fragment based screening. *Curr. Top. Med. Chem.* **2005**, *5* (8), 751−762.

(22) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432* (7019), 862−865.

(23) Verdonk, M. L.; Giangreco, I.; Hall, R. J.; Korb, O.; Mortenson, P. N.; Murray, C. W. Docking performance of fragments and druglike compounds. *J. Med. Chem.* **2011**, *54* (15), 5422−5431.

(24) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739−1749.

(25) Loving, K.; Salam, N. K.; Sherman, W. Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *J. Comput.-Aided Mol. Des.* **2009**, *23* (8), 541−554.

(26) Kawatkar, S.; Wang, H. M.; Czerminski, R.; Joseph-McCarthy, D. Virtual fragment screening: an exploration of various docking and scoring protocols for fragments using Glide. *J. Comput.-Aided Mol. Des.* **2009**, *23* (8), 527−539.

(27) Huang, Q.; Li, L. L.; Yang, S. Y. PhDD: A new pharmacophore-based de novo design method of drug-like molecules combined with assessment of synthetic accessibility. *J. Mol. Graph. Model.* **2010**, *28* (8), 775−787.

(28) Lippert, T.; Schulz-Gasch, T.; Roche, O.; Guba, W.; Rarey, M. De novo design by pharmacophore-based searches in fragment spaces. *J. Comput-Aided Mol. Des.* **2011**, *25* (10), 931−945.

(29) Besnard, J.; Ruda, G. F.; Setola, V.; Abecassis, K.; Rodriguiz, R. M.; Huang, X. P.; Norval, S.; Sassano, M. F.; Shin, A. I.; Webster, L. A.; Simeons, F. R.; Stojanovski, L.; Prat, A.; Seidah, N. G.; Constam, D. B.; Bickerton, G. R.; Read, K. D.; Wetsel, W. C.; Gilbert, I. H.; Roth, B. L.; Hopkins, A. L. Automated design of ligands to polypharmacological profiles. *Nature* **2012**, *492* (7428), 215−220.

(30) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. 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.* **1998**, *38* (3), 511−522.

(31) Degen, J.; Wegscheid-Gerlach, C.; Zaliani, A.; Rarey, M. On the art of compiling and using 'drug-like' chemical fragment spaces. *ChemMedChem* **2008**, *3* (10), 1503−1507.

(32) Böhm, H. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6* (1), 61−78.

(33) Nishibata, Y.; Itai, A. Automatic creation of drug candidate structures based on receptor structure - Starting point for artificial lead generation. *Tetrahedron* **1991**, *47* (43), 8985−8990.

(34) Mata, P.; Gillet, V.; Johnson, A.; Lampreia, J.; Myatt, G.; Sike, S.; Stebbings, A. SPROUT - 3D structure generation using templates. *J. Chem. Inf. Comput. Sci.* **1995**, 479−493.

(35) Stahl, M.; Todorov, N. P.; James, T.; Mauser, H.; Boehm, H. J.; Dean, P. M. A validation study on the practical use of automated de novo design. *J. Comput.-Aided Mol. Des.* **2002**, *16* (7), 459−478.

(36) Rotstein, S.; Murcko, M. Groupbuild - A fragment-based method for denovo drug design. *J. Med. Chem.* **1993**, *36* (12), 1700−1710.

(37) Wang, R.; Gao, Y.; Lai, L. LigBuilder: A multi-purpose program for structure-based drug design. *J. Mol. Model.* **2000**, 498−516.

(38) Schneider, G.; Lee, M.; Stahl, M.; Schneider, P. De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J. Comput.-Aided Mol. Des.* **2000**, *14* (5), 487−494.

(39) Fechner, U.; Schneider, G. Flux (1): A virtual synthesis scheme for fragment-based de novo design. *J. Chem. Inf. Model.* **2006**, *46* (2), 699−707.

(40) Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discovery* **2005**, *4* (8), 649−663.

(41) Vinkers, H.; de Jonge, M.; Daeyaert, F.; Heeres, J.; Koymans, L.; van Lenthe, J.; Lewi, P.; Timmerman, H.; Van Aken, K.; Janssen, P. SYNOPSIS: SYNthesize and OPtimize system in silico. *J. Med. Chem.* **2003**, *46* (13), 2765−2773.

(42) Hartenfeller, M.; Zettl, H.; Walter, M.; Rupp, M.; Reisen, F.; Proschak, E.; Weggen, S.; Stark, H.; Schneider, G. DOGS: Reaction-driven de novo design of bioactive compounds. *PLoS Comput Biol* **2012**, *8* (2), e1002380.

(43) Gillet, V.; Myatt, G.; Zsoldos, Z.; Johnson, A. SPROUT, HIPPO and CAESA: Tools for de novo structure generation and estimation of synthetic accessibility. *Perspect. Drug Discovery Des.* **1995**, *3*, 34−50.

(44) Boda, K.; Seidel, T.; Gasteiger, J. Structure and reaction based evaluation of synthetic accessibility. *J. Comput.-Aided Mol. Des.* **2007**, *21* (6), 311−325.

(45) Zaliani, A.; Boda, K.; Seidel, T.; Herwig, A.; Schwab, C. H.; Gasteiger, J.; Claussen, H.; Lemmen, C.; Degen, J.; Parn, J.; Rarey, M. Second-generation de novo design: A view from a medicinal chemist perspective. *J. Comput.-Aided Mol. Des.* **2009**, *23* (8), 593−602.

(46) Hartenfeller, M.; Eberle, M.; Meier, P.; Nieto-Oberhuber, C.; Altmann, K.; Schneider, G.; Jacoby, E.; Renner, S. A collection of robust organic synthesis reactions for in silico molecule design. *J. Chem. Inf. Model.* **2011**, *51* (12), 3093−3098.

(47) Miranker, A.; Karplus, M. Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins* **1991**, *11* (1), 29−34.

(48) Schubert, C. R.; Stultz, C. M. The multi-copy simultaneous search methodology: A fundamental tool for structure-based drug design. *J. Comput.-Aided Mol. Des.* **2009**, *23* (8), 475−489.

(49) Brenke, R.; Kozakov, D.; Chuang, G. Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25* (5), 621−627.

(50) Hall, D. R.; Ngan, C. H.; Zerbe, B. S.; Kozakov, D.; Vajda, S. Hot spot analysis for driving the development of hits into leads in fragment-based drug discovery. *J. Chem. Inf. Model.* **2012**, *52* (1), 199−209.

(51) Ngan, C. H.; Bohnuud, T.; Mottarella, S. E.; Beglov, D.; Villar, E. A.; Hall, D. R.; Kozakov, D.; Vajda, S. FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res.* **2012**, *40* (Web Server issue), W271−W275.

(52) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(53) Maass, P.; Schulz-Gasch, T.; Stahl, M.; Rarey, M. Recore: A fast and versatile method for scaffold hopping based on small molecule crystal structure conformations. *J. Chem. Inf. Model.* **2007**, *47* (2), 390−399.

(54) Thompson, D.; Denny, R.; Nilakantan, R.; Humblet, C.; Joseph-McCarthy, D.; Feyfant, E. CONFIRM: Connecting fragments found in receptor molecules. *J. Comput.-Aided Mol. Des.* **2008**, *22* (10), 761−772.

(55) Dey, F.; Caflisch, A. Fragment-based de novo ligand design by multiobjective evolutionary optimization. *J. Chem. Inf. Model.* **2008**, *48* (3), 679−690.

(56) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins* **1999**, *37* (1), 88−105.

(57) Gozalbes, R.; Carbajo, R. J.; Pineda-Lucena, A. Contributions of computational chemistry and biophysical techniques to fragment-based drug discovery. *Curr. Med. Chem.* **2010**, *17* (17), 1769−1794.

(58) ChemAxon. http://www.chemaxon.com (accessed April 4, 2013).

(59) Hoffer, L.; Chira, C.; Marcou, G.; Varnek, A.; Horvath, D. 2013, publication in preparation.

(60) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91* (1−3), 1−41.

(61) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.

(62) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Model.* **1998**, *38*, 983−996.

(63) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3−26.

(64) Irwin, J. J.; Shoichet, B. K. ZINC - A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45* (1), 177−182.

(65) Gasteiger, J.; Marsilli, M. A new model for calculating atomic charges in molecules. *Tetrahedron Lett.* **1978**, 3181−3184.

(66) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25* (2), 247−260.

(67) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossvry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; University of California, 2012.

(68) RCSB Protein Data Bank. http://www.rcsb.org/pdb/home/home.do (accessed April 4, 2013).

(69) Nazaré, M.; Matter, H.; Will, D. W.; Wagner, M.; Urmann, M.; Czech, J.; Schreuder, H.; Bauer, A.; Ritter, K.; Wehner, V. Fragment deconstruction of small, potent factor Xa inhibitors: exploring the superadditivity energetics of fragment linking in protein−ligand complexes. *Angew. Chem., Int. Ed. Engl.* **2012**, *51* (4), 905−911.

(70) Janin, Y. L. Heat shock protein 90 inhibitors. A text book example of medicinal chemistry? *J. Med. Chem.* **2005**, *48* (24), 7503−7512.

(71) Janin, Y. L. ATPase inhibitors of heat-shock protein 90, second season. *Drug Discovery Today* **2010**, *15* (9−10), 342−353.

(72) Barker, J. J.; Barker, O.; Boggio, R.; Chauhan, V.; Cheng, R. K. Y.; Corden, V.; Courtney, S. M.; Edwards, N.; Falque, V. M.; Fusar, F.; Gardiner, M.; Hamelin, E. M. N.; Hesterkamp, T.; Ichihara, O.; Jones, R. S.; Mather, O.; Mercurio, C.; Minucci, S.; Montalbetti, C.; Muller, A.; Patel, D.; Phillips, B. G.; Varasi, M.; Whittaker, M.; Winkler, D.; Yarnold, C. J. Fragment-based Identification of Hsp90 Inhibitors. *ChemMedChem* **2009**, *4* (6), 963−966.

(73) Barker, J. J.; Barker, O.; Courtney, S. M.; Gardiner, M.; Hesterkamp, T.; Ichihara, O.; Mather, O.; Montalbetti, C. A.; Müller, A.; Varasi, M.; Whittaker, M.; Yarnold, C. J. Discovery of a novel Hsp90 inhibitor by fragment linking. *ChemMedChem* **2010**, *5* (10), 1697−1700.

(74) Huth, J. R.; Park, C.; Petros, A. M.; Kunzer, A. R.; Wendt, M. D.; Wang, X. L.; Lynch, C. L.; Mack, J. C.; Swift, K. M.; Judge, R. A.; Chen, J.; Richardson, P. L.; Jin, S.; Tahir, S. K.; Matayoshi, E. D.; Dorwin, S. A.; Ladror, U. S.; Severin, J. M.; Walter, K. A.; Bartley, D. M.; Fesik, S. W.; Elmore, S. W.; Hajduk, P. J. Discovery and design of novel HSP90 inhibitors using multiple fragment-based design strategies. *Chem. Biol. Drug Des.* **2007**, *70*, 1−12.

(75) Murray, C. W.; Carr, M. G.; Callaghan, O.; Chessari, G.; Congreve, M.; Cowan, S.; Coyle, J. E.; Downham, R.; Figueroa, E.; Frederickson, M.; Graham, B.; McMenamin, R.; O'Brien, M. A.; Patel, S.; Phillips, T. R.; Williams, G.; Woodhead, A. J.; Woolford, A. J. Fragment-based drug discovery applied to Hsp90. Discovery of two lead series with high ligand efficiency. *J. Med. Chem.* **2010**, *53* (16), 5942−5955.

(76) Woodhead, A. J.; Angove, H.; Carr, M. G.; Chessari, G.; Congreve, M.; Coyle, J. E.; Cosme, J.; Graham, B.; Day, P. J.; Downham, R.; Fazal, L.; Feltell, R.; Figueroa, E.; Frederickson, M.; Lewis, J.; McMenamin, R.; Murray, C. W.; O'Brien, M. A.; Parra, L.; Patel, S.; Phillips, T.; Rees, D. C.; Rich, S.; Smith, D. M.; Trewartha, G.; Vinkovic, M.; Williams, B.; Woolford, A. J. Discovery of (2,4-dihydroxy-5-isopropylphenyl)-[5-(4-methylpiperazin-1-ylmethyl)-1,3-dihydroisoin-dol-2-yl]methanone (AT13387), a novel inhibitor of the molecular chaperone Hsp90 by fragment based drug design. *J. Med. Chem.* **2010**, *53* (16), 5956−5969.

(77) Sixma, T. K.; Smit, A. B. Acetylcholine binding protein (AChBP): A secreted glial protein that provides a high-resolution model for the extracellular domain of pentameric ligand-gated ion channels. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32* (1), 311−334.

(78) Edink, E.; Rucktooa, P.; Retra, K.; Akdemir, A.; Nahar, T.; Zuiderveld, O.; van Elk, R.; Janssen, E.; van Nierop, P.; van Muijlwijk-Koezen, J.; Smit, A. B.; Sixma, T. K.; Leurs, R.; de Esch, I. J. Fragment growing induces conformational changes in acetylcholine-binding protein: A structural and thermodynamic analysis. *J. Am. Chem. Soc.* **2011**, *133*, 5363−5371.

(79) Bemis, G.; Murcko, M. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39* (15), 2887−2893.

(80) McCourt, J. A.; Pang, S. S.; Guddat, L. W.; Duggleby, R. G. Elucidating the specificity of binding of sulfonylurea herbicides to acetohydroxyacid synthase. *Biochemistry* **2005**, *44* (7), 2330−2238.

(81) ChemAxon pKa Calculator Plugin. https://www.chemaxon.com/products/calculator-plugins/property-predictors/ (accessed February 2013).

(82) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T.; Mortenson, P. N.; Murray, C. W. Diverse, high-quality test set for the validation of protein−ligand docking performance. *J. Med. Chem.* **2007**, *50* (4), 726−741.

(83) Martin, E. J.; Sullivan, D. C. Surrogate AutoShim: Predocking into a universal ensemble kinase receptor for three dimensional activity prediction, very quickly, without a crystal structure. *J. Chem. Inf. Model.* **2008**, *48* (4), 873−881.

(84) Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **2007**, *47* (1), 195−207.

(85) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002.