Special Issue Article

Journal of Molecular Recognition

Received: 2 April 2009, Revised: 7 July 2009, Accepted: 15 July 2009, Published online in Wiley InterScience: 2009

(www.interscience.wiley.com) DOI:10.1002/jmr.981

Library screening by fragment-based docking

Danzhi Huang[a] and Amedeo Caflisch[a]*

We review our computational tools for high-throughput screening by fragment-based docking of large collections of small molecules. Applications to six different enzymes, four proteases, and two protein kinases, are presented. Remarkably, several *low-micromolar* inhibitors were discovered in each of the high-throughput docking campaigns. Probable reasons for the lack of *submicromolar* inhibitors are the tiny fraction of chemical space covered by the libraries of available compounds, as well as the approximations in the methods employed for scoring, and the use of a rigid conformation of the target protein. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: docking; screening; FFLD; SEED; DAIM; LIECE

INTRODUCTION

Among the many significant contributions to computational chemistry and physical chemistry, Martin Karplus has also pioneered the use of molecular fragments to map protein binding sites. His paper with Andrew Miranker on the simultaneous minimization of multiple copies of small and mainly rigid functional groups in the protein force field (MCSS) can be considered as the first fragment-based procedure for drug discovery.[1] Experimental techniques for fragment-based lead identification, such as structure–activity relationship by nuclear magnetic resonance spectroscopy,[2] share the same essential idea as the MCSS approach. Thus the importance of molecular fragments has been recognized and exploited first by computational approaches.[1,3,4] More recently, fragment-based drug discovery strategies have been developed using X-ray crystallography,[5] nuclear magnetic resonance spectroscopy,[6] surface plasmon resonance,[7] mass spectrometry,[8,9] substrate activity screening (where the fragments are substrates later converted into inhibitors[10–12]), and tethering.[13,14] Experimental techniques for fragment-based drug discovery have been discussed in previous reviews which contain a large number of applications.[15–20] Successful *in vitro* screening campaigns have been reported for several targets, and a non-exhaustive list includes β-secretase,[7,15,21,22] several protein kinases,[23–27] DNA gyrase,[28] caspase,[29,30] anthrax lethal factor,[31] and phosphodiesterase.[32]

In this review, we focus on the computational methods for fragment-based docking developed in our research group. We do not discuss *in silico* approaches developed by others (reviewed in[33–36]), but rather describe briefly our computational tools (in Section High-throughput fragment-based docking), and present applications (in Section *In silico* screening campaigns) to six different enzymes, four proteases and two protein kinases. Notably, it has been possible to identify single-digit micromolar inhibitors for all of the six enzymes by fragment-based docking of large libraries of small molecules. In the Conclusions section we give possible reasons for the difficulties encountered in discovering submicromolar inhibitors by *in silico* screening of collections of available compounds.

HIGH-THROUGHPUT FRAGMENT-BASED DOCKING

Two essential elements of our *in silico* screening approach are the fragment-based docking procedure and the scoring based on force field energy[37] with continuum electrostatics solvation.[38–42] The pipeline for high-throughput docking consists of four consecutive steps (Figure 1): (1) automatic decomposition of each molecule of the library into fragments,[43] (2) fragment docking and ranking of poses,[40,41] (3) flexible docking of each molecule of the library using the best poses of its fragments as anchors,[44] (4) evaluation of the binding free energy of multiple poses of each compound by the LIECE method (linear interaction energy (LIE) with continuum electrostatics[45]). The first three steps are performed by computer programs developed in our research group. In the fourth step, CHARMM[46,47] is used for the energy minimization and finite-difference Poisson calculations.

Decomposition of compounds into fragments

The automatic fragmentation of a molecule into substructures and the selection of the three anchor fragments for docking (see below) are performed by the program DAIM (Decomposition And Identification of Molecules[43]). The decomposition generates mainly rigid fragments which can be docked very efficiently (as explained in the next subsection). The decomposition of a molecule consists of four steps: ring identification, fragment definition, functional group merging, and completion of the valences. (i) Rings are identified by successively enumerating all neighbors (i.e., directly covalently bound atoms) of every atom,

* *Correspondence to: A. Caflisch, Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. E-mail: caflisch@bioc.uzh.ch*

a *D. Huang, A. Caflisch*
Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Abbreviations: *CDK2, cyclin-dependent kinase 2; DAIM, decomposition and identification of molecules; EphB4, ertythropoietin producing human hepatocellular carcinoma receptor tyrosine kinase B4; FFLD, fast flexible ligand docking; SEED, solvation energy for exhaustive docking.*

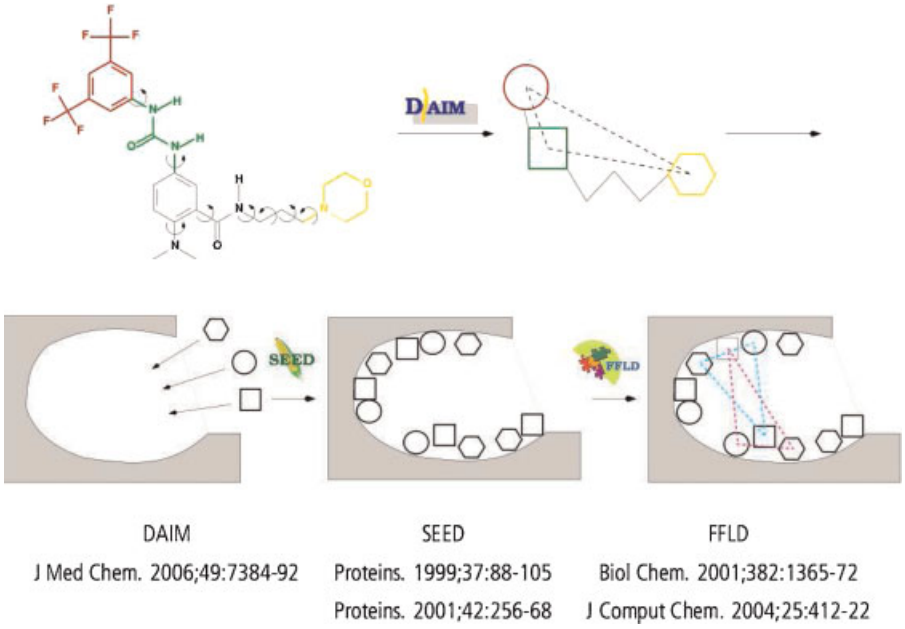

Figure 1. Schematic picture of the pipeline for automatic docking developed in our group. The automatic decomposition of the molecules into rigid fragments is carried out by the program DAIM.[43] The fragments are docked by the program SEED which evaluates the binding free energy including electrostatic solvation effects.[40] The program FFLD[44] is used for flexible ligand docking using positions and orientations of fragment triplets (dashed triangles).

similar to a breadth-first search. (ii) A fragment is defined as a set of atoms connected by unbreakable bonds. The basic definition of unbreakable bonds includes terminal, double, triple, aromatic bonds, and bonds in rings. Non-rotatable and unbreakable bonds are distinguished in DAIM; a non-rotatable bond is always unbreakable, whereas the reverse is not true (e.g., a double bond is non-rotatable and unbreakable, whereas an amide bond is unbreakable, but can assume more than one conformation). (iii) To form chemically relevant fragments and avoid the generation of many small groups, small functional groups (e.g., –OH, $–CH_3$, $–CX_3$ [where X can be any halogen], $–SO_3$, –CHO, $–NO_2$, $–NH_2$, and –SH) are merged with the fragment they are connected to. Unbreakable bonds and functional groups (points (ii) and (iii), respectively) can be defined by the user. (iv) In the final step, missing atom neighbors are added. An atom will lack a neighbor atom whenever the bond connecting them has been cut. These missing neighbors are replaced by hydrogen atoms to reconstitute the correct valence for every atom. A methyl group is used to fill valences where a hydrogen atom would result in an unwanted additional hydrogen bond direction (e.g., a hydrogen replacing a carbon atom bound to an sp^3 nitrogen).

Docking of anchor fragments

The docking approach implemented in the program SEED determines optimal positions and orientations of small to medium-size molecular fragments in the binding site of a protein.[40,41] Apolar fragments are docked into hydrophobic regions of the receptor while polar fragments are positioned such that at least one intermolecular hydrogen bond is formed. Each fragment is placed at several thousand different positions with multiple orientations (for a total of in the order of 10^6 poses) and the binding energy is estimated whenever severe clashes are not present (usually about 10^5 poses). The binding energy is the sum of the van der Waals interaction and the electrostatic energy. The latter consists of screened receptor-fragment interaction, as well as receptor and fragment desolvations calculated by an efficient numerical implementation of the generalized Born approach.[39]

Flexible docking of library compounds

The flexible-ligand docking approach FFLD uses a genetic algorithm and a very efficient scoring function.[44,48] The random perturbations (i.e., sampling) in the genetic algorithm affect only the conformation of the ligand; its placement in the binding site is determined by the SEED anchors and a least square fitting method.[49] In this way the position and orientation of the ligand in the binding site are determined by the best binding modes of its fragments previously docked by SEED. On the other hand, the scoring function used in FFLD is based on van der Waals and hydrogen bond terms and does not explicitly include solvation for efficiency reasons. Solvation effects are implicitly accounted for because the poses of the fragments are previously sorted according to force field energy with electrostatic solvation in SEED. FFLD requires three not-necessarily different fragments to place a flexible ligand unambiguously in the binding site, e.g., the fluorobenzene, morpholine, and benzoic acid of compound **2** (Table 1). The automatic definition of the three anchor fragments for each molecule of the library is performed by DAIM using the chemical richness, i.e., the sum of all entries in the DAIM fingerprint.[43]

LIECE binding energy evaluation

The essential idea of linear interaction energy (LIE) models is that the free energy of complex formation can be calculated by considering only the end points of the thermodynamic cycle of

Table 1. The library screening campaigns by fragment-based docking performed at the University of Zurich

		Most potent inhibitor							
Protein	Enzyme class	Cpd.	Affinity (μM)	Ligand efficiency (kcal/mol per non-hydrogen atom)	MW (g/mol)	Hit rate[a]	Size of library	Scoring method	Reference
Proteases									
β-secretase	Asp PR	**1**	3.0[b]	0.19	561	17% (12/72)	512000	LIECE	83
β-secretase	Asp PR	**2**	7.1[c]	0.19	538	10% (9/88)	316000	LIECE	84
Plasmepsin	Asp PR	**3**	2[c]	0.25	500	32% (6/19)	40000	Consensus scoring	63
NS3 protease	Ser PR	**4**	40[d]	0.33	282	5% (1/22)	12000	LIECE	64
NS3 protease	Ser PR	**5**	2.8[c]/90[d]	0.34/0.25	298	40% (2/5)	19000	Pose filtering	98
Cathepsin B	Cys PR	**6**	4.8/6.7[c]	0.29/0.28	369	3% (1/29)	48000	Consensus scoring	113
Kinases									
EphB4	Tyr kin.	**7**	1.5[c]	0.32	337	19% (8/43)	728000	LIECE	115
CDK2	Ser/Thr kin.	**8**	7.8[c]	0.32	321	3% (1/30)	40000	LIECE	65

[a] The ratio in parentheses is the number of compounds with a value of the measured affinity below 100 μM divided by the number of molecules tested experimentally.
[b] Cell-based assay.
[c] Enzymatic assay with purified protein in solution.
[d] Binding affinity measured by nuclear magnetic resonance spectroscopy.

ligand binding, i.e., bound and free states. For this purpose, one usually calculates average values of interaction energies from molecular dynamics simulations of the isolated ligand and the ligand/protein complex.[50,51] The free energy of binding is approximated by

$$\Delta G = \alpha\left(\langle E^{\mathrm{vdW}}\rangle_{\mathrm{bound}} - \langle E^{\mathrm{vdW}}\rangle_{\mathrm{free}}\right) + \beta\left(\langle E^{\mathrm{elec}}\rangle_{\mathrm{bound}} - \langle E^{\mathrm{elec}}\rangle_{\mathrm{free}}\right) \quad (1)$$

where E^{vdW} and E^{elec} are the van der Waals and electrostatic interaction energies between the ligand and its environment. The environment is either the solvent (free) or both the protein and solvent (bound). The $\langle\rangle$ denotes an ensemble average sampled over a molecular dynamics[50] or Monte Carlo[52] trajectory. The coefficient α is determined empirically.[50] Originally, β was fixed to a value of 1/2, as predicted by the linear response approximation.[50] Later studies have shown, however, that improved models for a large variety of systems could be obtained by considering β as a free parameter.[53] Consequently, both coefficients are obtained by a fit of experimentally determined values of ΔG to the calculated values of E^{elec} and E^{vdW} for a training set of known ligands.

The original LIE method and modifications thereof have been applied to a large number of existing inhibitor/protein datasets.[50,54–59] Moreover, LIE-based scorings of ligands were shown to perform better than established scoring functions.[60] Interestingly, recent applications to pharmaceutically relevant enzyme targets have documented the predictive ability and usefulness in lead-discovery projects. As an example, the LIE method with explicit water molecular dynamics sampling was successfully used in the design of a series of inhibitors of the malarial aspartic proteases plasmepsin I and II.[61] Unfortunately, LIE cannot be used for high-throughput docking because of its computational requirements (the currently fastest implementation needs about 6 h for each compound[60]). Therefore, we have replaced the explicit water molecular dynamics (or Monte Carlo) simulations with a simple energy minimization and combined the LIE method with a rigorous treatment of solvation within the continuum electrostatics approximation,[45] i.e., the numerical solution of the Poisson equation by the finite-difference technique.[62] The LIECE approach, where the last two letters stand for continuum electrostatics, is about two orders of magnitude faster than previous LIE methods and shows a similar precision on the targets tested. In fact, we have observed an error of about 1 kcal/mol for 13 and 29 peptidic inhibitors of β-secretase and HIV-1 protease, respectively.[45] Similar accuracy has been reported for other proteases[63,64] and five kinases.[65]

Consensus scoring

It has been reported that consensus scoring is generally preferable to the use of a single scoring function.[66] Furthermore, the median rank is more suitable than the average rank in consensus-scoring because the former is less sensitive to outliers.[67] Rank by median consensus scoring was used in the *in silico* screening against plasmepsin and cathepsin B (Table 1). For plasmepsin, consensus scoring was preferred to ranking by LIECE because visual inspection of the best LIECE poses revealed several unlikely binding modes. In the case of cathepsin B, the lack of experimental data on inhibitors binding outside the catalytic center was the reason for not using LIECE which requires at least 10–15 binding affinity data points for fitting the 2–3 parameters of the linear model.

IN SILICO SCREENING CAMPAIGNS

During 2004–2008, our suite of programs for fragment-based docking[40,43,44] has been employed in eight *in silico* screening

campaigns on six different enzymes which play a key role in a variety of diseases (Table 1). Four of these enzymes are proteases of three different classes (aspartic, serine, and cysteine proteases), while the remaining two are a tyrosine kinase and a Ser/Thr kinase. The 2D structures of the most potent inhibitors discovered in the eight high-throughput docking campaigns are shown in Figure 2.

β-Secretase (Alzheimer's disease)

Alzheimer's disease is the most common neurodegenerative disease and it accounts for the majority of the cases of dementia diagnosed after the age of 60.[68] Amyloid plaques, which are found in the post-mortem brain of Alzheimer's disease patients,[69,70] consist mainly of fibrillar aggregates of the Aβ peptide, a proteolytic cleavage product of the β-amyloid precursor protein (APP). Two enzymes, γ- and β-secretase (β-site APP cleaving enzyme) are responsible for the sequential processing of APP.[71] Although it is not clear whether the plaques or oligomeric prefibrillar species are responsible for neuronal loss and dementia,[72] the pepsin-like aspartic protease β-secretase has become one of the major Alzheimer's disease targets.[68,73–77] β-Secretase inhibitors have been shown to lower brain Aβ after direct intracranial administration[76,78] or via peripheral administration at relatively high doses in murine models.[75,76] Moreover, a novel potent tertiary carbinamine inhibitor of β-secretase effectively lowers Aβ levels in a non-human primate model.[77]

β-Secretase is not an easy target to block.[68,73,79–81] For instance, only a single molecule (1,3,5-trisubstituted benzene) emerged as β-secretase inhibitor from a multimillion compound library submitted to a high-throughput screening campaign.[82] As a proof-of-principle of our *in silico* screening approach, high-throughput fragment-based docking into the β-secretase active site and LIECE binding free energy evaluation has led to the discovery of three novel series of inhibitors: phenylurea derivatives (e.g., compound **1**),[83] triazine derivatives (e.g.,

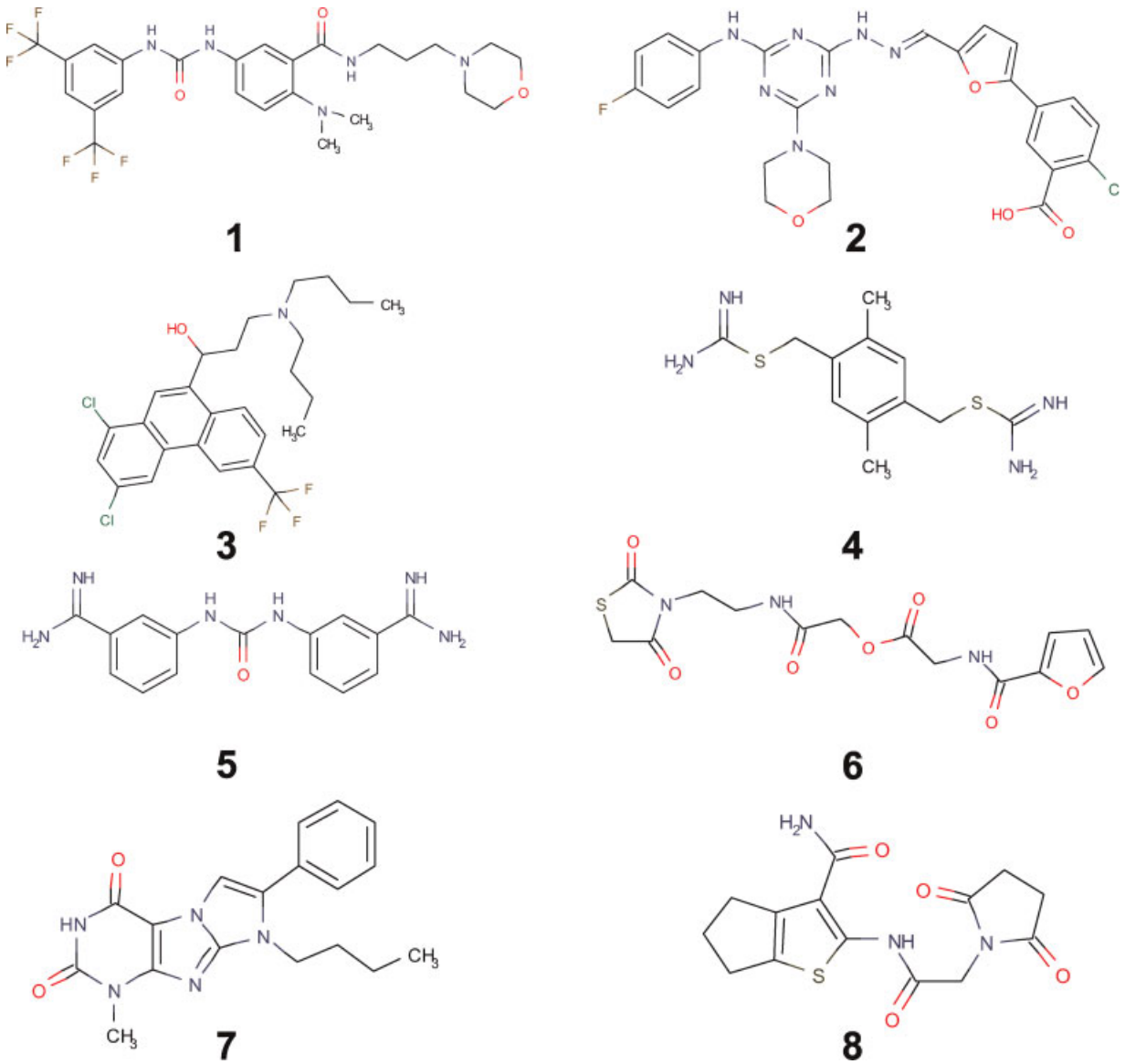

Figure 2. Low micromolar inhibitors discovered by high-throughput fragment-based docking into β-secretase (compounds **1** and **2**), plasmepsin II (**3**), West Nile virus NS3 protease (**4** and **5**), cathepsin B (**6**), EphB4 tyrosine kinase (**7**), and CDK2 Ser/Thr kinase (**8**). See Figure 1 for the programs used for docking, and Table 1 for hit rates, sizes of libraries screened, and experimentally measured affinities.

compound **2**),[84] and a set of five cell-permeable, non-peptide, low-micromolar inhibitors with a different scaffold (D. Huang and A. Caflisch, unpublished results). Among them, the phenylurea derivatives were identified from an initial set of about half a million molecules[83] (Table 1). Twelve of the 72 tested compounds inhibit β-secretase in at least one of two different mammalian cell-based assays ($EC_{50} < 10\ \mu M$). It is important to note that for almost all of the 12 compounds, for which an EC_{50} value could be measured, the discrepancies between LIECE-predicted affinity and the experimental value is within the LIECE accuracy of about 1 kcal/mol.[45] The triazine derivatives were selected from an initial set of about 300 000.[84] Ten of the 88 tested compounds inhibit β-secretase activity in an enzymatic assay ($IC_{50} < 100\ \mu M$), and four of them are active in a mammalian cell-based assay ($EC_{50} < 20\ \mu M$).

P. *falciparum* plasmepsin II (malaria)

An estimated 40% of the world's population is a potential victim of malaria, which is responsible for 300–660 million infections annually.[85] Furthermore, there is an urgent need to develop new antimalarial medicines because of the emerging drug resistance.[86] Plasmepsins are pepstatin-like aspartic proteases unique to the malaria parasites (which belong to the genus *Plasmodium*). They are involved in metabolism and host cell invasion.[87] Ten plasmepsins have been identified in the genome of *P. falciparum*, the plasmodium species that causes the most fatal form of malaria in human, and four of these 10 are located in the *food vacuole*, an acidic lysosome-like organelle in which hemoglobin degradation takes place. It has been shown that inhibitors of plasmepsins are fatal to the parasites,[88] which suggests that plasmepsins are pharmaceutically relevant targets. Furthermore, several small-molecule inhibitors of retroviral and human aspartic proteases, namely HIV-protease[89] and renin,[90] are effective and safe medicines, which provides additional support to the relevance of plasmepsins as drug targets.

We used our fragment-based docking procedure to search for inhibitors of plasmepsin II.[63] A total of 4.6 million compounds were first clustered according to 2D structural similarity resulting in about 40 000 molecules which were then used for fragment-based docking. Docking into the plasmepsin II active site was followed by consensus scoring using four force field-based energy functions. A total of 19 compounds were tested in an enzymatic assay, and three of them showed single-digit micromolar inhibitory activity (Table 1).[63] One of these three inhibitors is halofantrine (compound **3**), an antimalarial drug discovered more than 40 years ago whose mechanism of action is still unknown. To better investigate the binding mode of halofantrine, four 50 ns molecular dynamics simulations with explicit solvent were performed starting from two different poses, one generated by automatic docking and the other by manual fitting with the help of a computer graphics program. The molecular dynamics simulations indicate that the binding mode generated by fragment-based docking is more stable than the one obtained by manual docking although it is not possible to definitively discard either.[63]

West Nile virus NS3 protease (flaviviral infections)

The pathogenic members of the flavivirus family, e.g., West Nile virus and the closely related Dengue virus, are transmitted by mosquito bites. Although an estimated 2.5 billion people are potential victims of encephalitis and other fatal maladies caused by flaviviruses,[91] these diseases have received much less attention than other tropical diseases like avian influenza. Their status as "neglected" diseases is in part due to the fact that flaviviruses are widespread mainly in poor countries, and mosquitos can fly only much shorter distances than migratory birds so that they do not represent a threat in developed countries. Recently, the non-structural 3 (NS3) protease has been shown to be responsible for cleavage of the viral polyprotein precursor and to play a pivotal role in the replication of flaviviruses.[92,93] In fact, site directed mutagenesis focused on the NS3 protease cleavage sites in the polyprotein precursor abolishes viral infectivity.[93] Therefore, the NS3 protease is one of the most promising targets for drug development against flaviviridae infections. In this context it is important to note that two inhibitors of the closely related hepatitis C virus protease are under late-stage clinical development.[94–97]

We have run two *in silico* screening campaigns to identify inhibitors of the West Nile virus NS3 protease. The first high-throughput docking campaign (Figure 3, left) was performed on the X-ray structure of the protease[64] while the second campaign (Figure 3, right) made use of a snapshot selected along a 1 ns explicit solvent molecular dynamics simulation started from the crystal structure.[98] This snapshot was chosen from a set of 100 as it optimally accommodates three representative molecular fragments: benzene and two functional groups with a positive charge. The former is the most common ring in the known drugs, while the latters were employed because the active site of the NS3 protease has a large amount of hydrogen bond acceptors. Interestingly, the hit rate was high in both campaigns (5 and 40%, Table 1). Most importantly, inhibitors **4** and **5** are candidate lead compounds. They occupy only the S1 and S2 pockets of the substrate binding site (Figure 4), so that additional substituents are expected to improve their affinity and, at the same time, retain good druglike properties as they both have very favorable ligand efficiency (which is defined as the experimentally measured free energy of binding divided by the number of non-hydrogen atoms[99]). The *in silico* discovery of compounds **4** and **5** is remarkable because the West Nile virus NS3 protease is a very difficult target. Very few non-peptidic inhibitors have been reported. Moreover, less than 10 molecules emerged as inhibitors (in the micromolar range) from a library of more than one million compounds submitted to a high-throughput *in vitro* screening campaign.[100]

It is interesting to note that compound **5**, as well as another inhibitor of the West Nile virus NS3 protease discovered by docking into the molecular dynamics snapshot (the diphenylesther **2** in Reference[98]), would not have been identified by docking into the X-ray structure, as the S1 pocket in the latter does not accommodate the benzene ring in an energetically favorable way.[98]

Cathepsin B (cancer and rheumatic disorders)

Cathepsin B is capable of endopeptidase,[101] peptidyl-dipeptidase,[102,103] and carboxypeptidase activities.[104,105] Among the cysteine peptidases cathepsin B is unique for the presence of a flexible segment, known as the occluding loop, that can block the primed subsites of the substrate binding cleft. With the occluding loop in the open conformation cathepsin B acts as an endopeptidase, while it acts as an exopeptidase when the loop is closed. Cathepsin B is involved in a number of human disorders. It

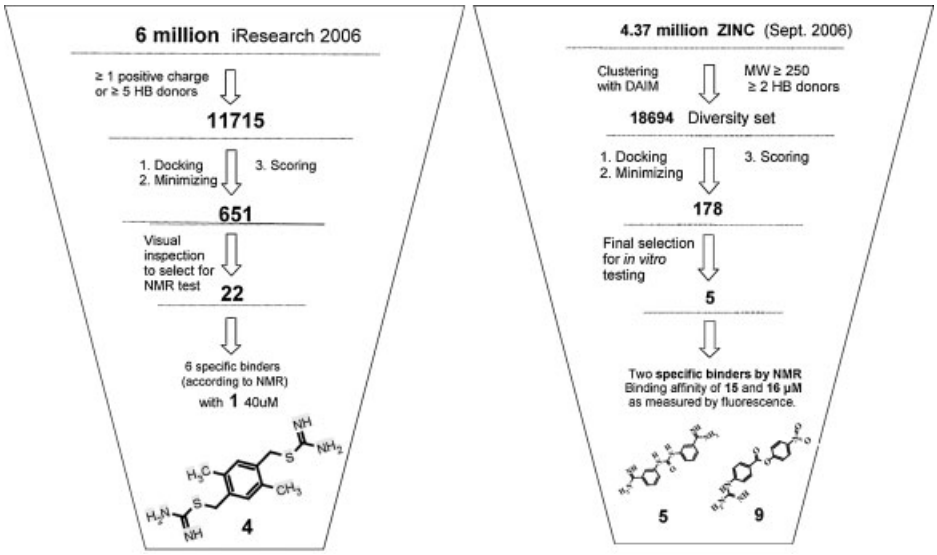

Figure 3. Schematic picture of the two *in silico* screening campaigns against West Nile virus NS3 protease. Docking of the compounds was performed by DAIM/SEED/FFLD using the 2fp7 structure of the WNV protease as explained in the text and References.[64,98]

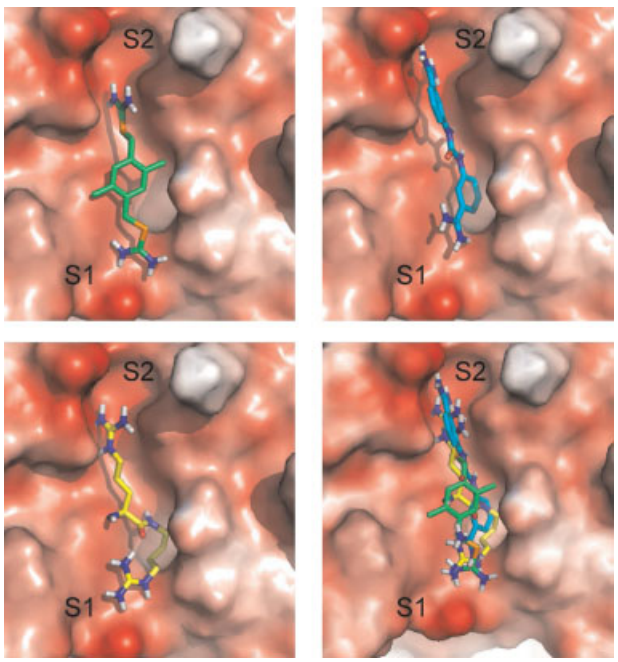

Figure 4. The guanidinium groups of inhibitors **4** (top left) and **5** (top right) are involved in electrostatic interactions with hydrogen bond acceptors in the S1 and S2 pockets of the West Nile virus NS3 protease, as observed in the X-ray structure of the aldehyde peptidic inhibitor benzoyl-Nle-Lys-Arg-Arg-H (PDB code 2fp7) (bottom left). Only the C-terminal dipeptide Arg-Arg-H of the aldehyde inhibitor is shown for clarity. (Bottom right) Overlap of the poses of compounds **4** and **5**, which were generated by fragment-based docking, with the binding mode of the peptidic inhibitor. The inhibitors are shown by sticks colored by atom-type with carbon atoms in green, cyan, and yellow for compound **4**, **5**, and the peptidic inhibitor, respectively. The surface of WNV NS2B-NS3pro is colored by electrostatic potential with red and blue for negative and positive potential, respectively. The figure was prepared using PyMOL (Delano Scientific, San Carlos, CA) and the APBS program[126] was used for calculation of the electrostatic surface.

activates trypsinogen in hereditary pancreatitis,[106] participates in apoptosis,[107] tumor progression and malignancy,[108,109] and plays an important role in rheumatic diseases.[110–112]

We have targeted the occluding loop of human cathepsin B outside the catalytic center, using high-throughput fragment-based docking.[113] The aim was to identify inhibitors that would interact with the occluding loop thereby modulating enzyme activity without the help of chemical warheads (i.e., without reactive functional groups that form a covalent bond with residues in the catalytic site but usually also bind unspecifically to other proteins). From a library of about 48 000 compounds, the *in silico* approach identified compound **6** which fulfills the working hypothesis (Table 1). This molecule possesses two distinct binding moieties and behaves as a reversible, double-headed competitive inhibitor of cathepsin B by excluding synthetic and protein substrates from the active center. The kinetic mechanism of inhibition suggests that the occluding loop is stabilized in its closed conformation, mainly by hydrogen bonds with the inhibitor, thus decreasing endoproteolytic activity of the enzyme. Furthermore, the dioxothiazolidine head of compound **6** sterically hinders binding of the C-terminal residue of substrates resulting in inhibition of the exopeptidase activity of cathepsin B in a physiopathologically relevant pH range.[113]

EphB4 tyrosine kinase (cancer)

The protein kinase EphB4 (ertythropoietin producing human hepatocellular carcinoma receptor tyrosine kinase B4) is a highly attractive angiogenic target involved in many types of cancer.[114] It seems to be rather recalcitrant to inhibition because, despite its potential therapeutic importance, very few inhibitors have been reported in the literature up to date.

We have developed and applied the ALTA (anchor-based library tailoring) approach to identify ATP-competitive inhibitors of EphB4[115] (Figure 5). ALTA is an automatic fragment-based procedure for focusing libraries of compounds. First, molecular fragments are docked and prioritized (i.e., anchors are selected

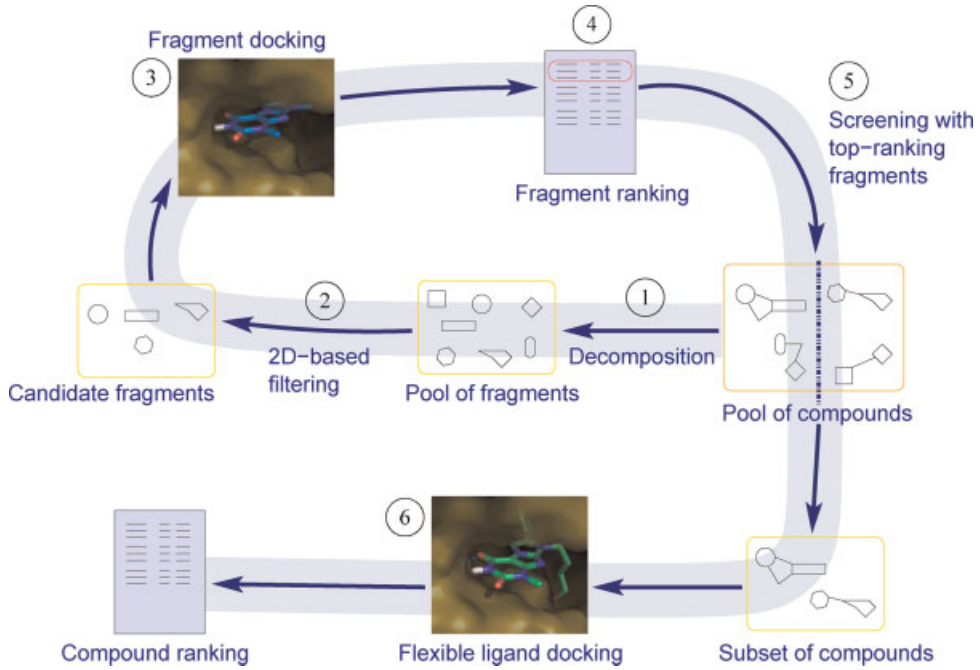

Figure 5. Graphical representation of the workflow of the ALTA procedure.[115] The first step is the automatic decomposition[43] of a library of compounds (orange rectangle) to obtain the pool of fragments. Afterwards, fragments selected based on the binding site features are docked[40] and ranked according to their binding energy.[41] Poses for molecules that contain at least one of the top-ranking fragments are then generated by flexible-ligand docking.[44]

based on a ranking) according to force field energy which includes continuum electrostatics solvation.[40,39] Large collections of molecules can then be effectively reduced in size by selecting only the compounds that have one (or more) fragment among the top ranking anchors. In principle, ALTA does not require any information about known inhibitors but in the application to the EphB4 kinase pharmacophore knowledge (hydrogen bonds to the hinge region) was additionally used to efficiently reduce the size of the original library from about 728 000 to 21 418 molecules. Two series of novel EphB4 inhibitors have been identified by ALTA. Compound **7** is a potential candidate for further development because of its low-micromolar affinity and molecular weight of only 337 DA which result in a very favorable ligand efficiency (Table 1). Moreover, the kinetic characterization of a very similar compound (**2** in Table 1 of Reference[115]) indicates that this series of molecules bind to the ATP-binding site, as predicted by the docking calculations.[115] Recently, single-digit nanomolar affinities have been reached by chemical synthesis of derivatives of compound **7** that were designed on the basis of the binding mode obtained by docking.[116]

CDK2 Ser/Thr kinase (cancer)

The human cyclin-dependent kinase 2 (CDK2) is a Ser/Thr kinase that controls cell cycle progression in proliferating eukaryotic cell.[117–119] The activity of CDK2 is tightly controlled, and fully activated CDK2 is essential for proper S phase progression. Studies have shown that inhibition of CDK2/cyclin A during S phase leads to S phase arrest and apoptosis, which has suggested a pharmacological role for CDK2 inhibitors in the treatment of cancer.[120]

A diversity set of 40 375 compounds was used for fragment-based docking into the X-ray structure of CDK2 (PDB code 1KE5). A threshold in the ratio between van der Waals energy and molecular weight, and the presence of at least one key hydrogen bond were used as filters to discard unfavorable poses. The remaining poses were scored by the two-parameter LIECE model fitted on 73 known CDK2 inhibitors. Thirty compounds were tested in an enzymatic assay, and compound **8** emerged with a single-digit micromolar IC_{50} value and a favorable ligand efficiency (0.32 kcal/mol per non-hydrogen atom, Table 1).[65]

CONCLUSIONS

Low micromolar inhibitors of four proteases and two protein kinases have been identified by high-throughput screening using fragment-based docking. The catalytic sites of the proteases have very different shape from the ATP-binding sites of the protein kinases. Moreover, the former show a broad range of surface-hydrophilicity ranging from mainly hydrophobic (β-secretase) to a strong electrostatic potential (NS3 protease). It is thus encouraging that docking finds inhibitors for all of them. The six enzymes are involved in key biological processes in humans or human parasites. Therefore, they are relevant drug targets: β-secretase in Alzheimer's disease, cathepsin B in cancer and rheumatic disorders, EphB4 tyrosine kinase in cancer-related angiogenesis, CDK2 Ser/Thr kinase in several types of cancer, plasmepsin II in malaria, and NS3 protease in infections caused by flaviviruses, in particular West Nile virus and Dengue virus. Recently, a series of 50 derivatives of the inhibitor **7** of the EphB4 tyrosine kinase have been synthesized. These 50 compounds were designed using the binding mode obtained by automatic docking. Notably, six of these compounds are active in the submicromolar range and one of them has an IC_{50} value of about 5 nM.[116] It is likely that medicinal chemistry optimization of all of the inhibitors presented in this review article (i.e., discovered by high-throughput docking) might yield (low) nanomolar inhibitors but it is very difficult to predict in advance the eventual improvement in potency.

One essential component of our scoring approach is the use of a force field energy supplemented by the evaluation of electrostatic solvation effects. The latter are calculated by models of aqueous solvent based on the continuum dielectric approximation. For fragment docking, the generalized Born approximation is used for evaluating the electrostatic component of the binding free energy which includes screened electrostatic interaction, and protein and fragment desolvation terms.[40,41] The finite-difference Poisson equation is employed to calculate electrostatic solvation effects for the poses of the compounds obtained by flexible ligand docking. The poses are then usually ranked by the LIECE approach.[45,65]

The two main outcomes of our *in silico* screening campaigns on six different enzymes are that the hit rate is rather high (between 3 and 40%) and the most potent inhibitors have low micromolar affinity and favorable ligand efficiency (Table 1). It is difficult to compare hit rates reported in different studies due to differences in the libraries employed for screening and in the affinity thresholds adopted for the definition of hits. Yet, our hit rates compare favorably with those reported for experimental high-throughput screening (0.01–0.1%) and fragment-based screening (0.1–1%) techniques. Moreover, *in silico* screening involves much smaller time, consumables, and labor costs than *in vitro* screening.

It was not possible to identify submicromolar inhibitors in the eight high-throughput docking campaigns (for six different enzymes), probably because one or more of the three following reasons. First, the libraries of available compounds cover a very small fraction of the potential druglike molecules. In fact, it has been estimated that the number of potential fragments with up to 11 heavy atoms (under constraints due to chemical stability and synthetic feasibility) is on the order of 10^7,[121] while the number of druglike molecules with less than 30 heavy atoms is larger than 10^{60}.[122] Furthermore, it is likely that the coverage of chemical space is very heterogeneous in the libraries of compounds that are available. A recent analysis of commercially available compounds suggests that that the chemical space is skewed mainly toward ligands of G-protein coupled receptors (10.6% of the compounds in a database of about 1 million small molecules resemble known ligands of these receptors) with fewer kinase-like ligands (4.2%) and an even smaller amount of compounds resembling protease inhibitors (2.3%).[123] Note also that an 80-nM ATP-competitive inhibitor of casein kinase II has been identified by high-throughput docking of a subset of the Novartis collection of compounds (about 400 000 molecules in 2003, which are not in the public domain).[124]

Second, there are significant sources of errors in both the force field and continuum dielectric model used to approximate solvation effects. As an example polarization effects are neglected in force fields based on fixed partial charges. A recent extension of the LIECE model has emphasized the importance of quantum mechanics to capture these effects, in particular for sets of inhibitors with significant differences in the number of formal charges.[125]

Third, all docking studies presented in this review were performed with a rigid protein. The rigid-protein approximation, which is required for efficiency reasons, dramatically restricts the number of favorable poses and results in false negatives. As an example, two small-molecule inhibitors of the NS3 protease have been identified by docking into a molecular dynamics snapshot, which would not have been possible by using the X-ray structure.[98]

In conclusion, we have reviewed our computational tools for fragment-based library docking and presented applications to six different enzymes. The scoring of poses based on force field energy with implicit solvation has shown robustness in identifying low micromolar inhibitors. Therefore, for target proteins of known three-dimensional structure, the efficiency and high hit rate of our fragment-based docking approach makes it a cost-effective alternative to experimental screening techniques, in particular in the lead identification phase of the drug discovery process.

Acknowledgements

A.C. thanks the talented and dedicated PhD students and post-docs who have developed and applied the methods for fragment-based docking and scoring. The computer programs have been developed mainly by three PhD students: N. Budin (FFLD), P. Kolb (DAIM), and N. Majeux (SEED). The *in silico* screening campaigns have been carried out by P. Alfarano (cathepsin B), M. Convertino (β-Secretase), D. Ekonomiuk (NS3 protease), R. Friedman (plasmepsin II), and P. Kolb (kinases). Very generous financial support has been provided by the innovation promotion agency (KTI) of the Swiss government, and the Canton of Zürich. All of our computer programs for fragment-based docking are available for free at http://www.biochem-caflisch.uzh.ch/download/

REFERENCES

1. Miranker A, Karplus M. 1991. Functionality maps of binding sites: a multiple copy simultaneous search method. *Protein. Struct. Funct. Genet.* **11**: 29–34.
2. Shuker H, Hajduk P, Meadows R, Fesik SW. 1996. Discovering high affinity ligands for proteins: SAR by NMR. *Science* **274**: 1531–1534.
3. Boehm HJ. 1992. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided. Mol. Des.* **6**(1): 61–78.
4. Caflisch A, Miranker A, Karplus M. 1993. Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *J. Med. Chem.* **36**: 2142–2167.
5. Nienaber VL, Richardson PL, Klighofer V, Bouska JJ, Giranda VL, Greer J. 2000. Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotechnol.* **18**: 1105–1108.
6. Pellecchia M, Bertini I, Cowburn D, Dalvit C, Giralt E, Jahnke W, James TL, Homans SW, Kessler H, Luchinat C, Meyer B, Oschkinat H, Peng J, Schwalbe H, Siegal G. 2008. Perspectives on NMR in drug discovery: a technique comes of age. *Nat. Rev. Drug. Discov.* **7**(9): 738–745.
7. Geschwindner S, Olsson L-L, Albert JS, Deinum J, Edwards PD, de Beer T, Folmer RHA. 2007. Discovery of a novel warhead against beta-secretase through fragment-based lead generation. *J. Med. Chem.* **50**(24): 5903–5911.
8. Swayze EE, Jefferson EA, Sannes-Lowery KA, Blyn LB, Risen LM, Arakawa S, Osgood SA, Hofstadler SA, Griffey RH. 2002. SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *J. Med. Chem.* **45**: 3816–3819.
9. Ockey DA, Dotson JL, Struble ME, Stults JT, Bourell JH, Clark KR, Gadek TR. 2004. Structure-activity relationships by mass spectrom-

etry: identification of novel MMP-3 inhibitors. *Bioorg. Med. Chem.* **12**: 37–44.
10. Wood WJL, Patterson AW, Tsuruoka H, Jain RK, Ellman JA. 2005. Substrate activity screening: a fragment-based method for the rapid identification of nonpeptidic protease inhibitors. *J. Am. Chem. Soc.* **127**(44): 15521–15527.
11. Patterson AW, Wood WJL, Hornsby M, Lesley S, Spraggon G, Ellman JA. 2006. Identification of selective, nonpeptidic nitrile inhibitors of cathepsin s using the substrate activity screening method. *J. Med. Chem.* **49**(21): 6298–6307.
12. Patterson AW, Wood WJL, Ellman JA. 2007. Substrate activity screening (SAS): a general procedure for the preparation and screening of a fragment-based non-peptidic protease substrate library for inhibitor discovery. *Nat. Protoc.* **2**(2): 424–433.
13. Erlanson DA, Lam JW, Wiesmann C, Luong TN, Simmons RL, DeLano WL, Choong IC, Burdett MT, Flanagan WM, Lee D, Gordon EM, O'Brien T. 2003. In situ assembly of enzyme inhibitors using extended tethering. *Nat. Biotechnol.* **21**: 308–314.
14. Erlanson DA, Wells JA, Braisted AC. 2004. Tethering: fragment-based drug discovery. *Annu. Rev. Biophys. Biomol. Struct.* **33**: 199–223.
15. Hajduk PJ, Greer J. A decade of fragment-based drug design: strategic advances and lessons learned. 2007. *Nat. Rev. Drug. Discov.* **6**(3): 211–219.
16. Jhoti H, Cleasby A, Verdonk M, Williams G. 2007. Fragment-based screening using X-ray crystallography and NMR spectroscopy. *Curr. Opin. Chem. Biol.* **11**: 485–493.
17. Siegal G, Ab E, Schultz J. 2007. Integration of fragment screening and library design. *Drug. Discov. Today.* **12**: 1032–1039.
18. Congreve M, Chessari G, Tisi D, Woodhead AJ. 2008. Recent developments in fragment-based drug discovery. *J. Med. Chem.* **51**(13): 3661–3680.
19. Fattori D, Squarcia A, Bartoli S. 2008. Fragment-based approach to drug lead discovery: overview and advances in various techniques. *Drugs R D* **9**: 217–227.
20. Hesterkamp T, Whittaker M. 2008. Fragment-based activity space: smaller is better. *Curr. Opin. Chem. Biol.* **12**: 260–268.
21. Murray CW, Callaghan O, Chessari G, Cleasby A, Congreve M, Frederickson M, Hartshorn MJ, McMenamin R, Patel S, Wallis N. 2007. Application of fragment screening by X-ray crystallography to beta-secretase. *J. Med. Chem.* **50**: 1116–1123.
22. Congreve M, Aharony D, Albert J, Callaghan O, Campbell J, Carr RAE, Chessari G, Cowan S, Edwards PD, Frederickson M, McMenamin R, Murray CW, Patel S, Wallis N. 2007. Application of fragment screening by X-ray crystallography to the discovery of aminopyridines as inhibitors of beta-secretase. *J. Med. Chem.* **50**: 1124–1132.
23. Furet P, Bold G, Hofmann F, Manley P, Meyer T, Altmann K-H. 2003. Identification of a new chemical class of potent angiogenesis inhibitors based on conformational considerations and database searching. *Bioorg. Med. Chem. Lett.* **13**(18): 2967–2971.
24. Gill A. 2004. New lead generation strategies for protein kinase inhibitors - fragment based screening approaches. *Mini. Rev. Med. Chem.* **4**(3): 301–311.
25. Szczepankiewicz BG, Liu G, Hajduk PJ, Abad-Zapatero C, Pei Z, Xin Z, Lubben TH, Trevillyan JM, Stashko MA, Ballaron SJ, Liang H, Huang F, Hutchins CW, Fesik SW, Jirousek MR. 2003. Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *J. Am. Chem. Soc.* **125**: 4087–4096.
26. Warner SL, Bashyam S, Vankayalapati H, Bearss DJ, Han H, Mahadevan D, Hoff DDV, Hurley LH. 2006. Identification of a lead small-molecule inhibitor of the Aurora kinases using a structure-assisted, fragment-based approach. *Mol. Cancer. Ther.* **5**: 1764–1773.
27. Wyatt PG, Woodhead AJ, Berdini V, Boulstridge JA, Carr MG, Cross DM, Davis DJ, Devine LA, Early TR, Feltell RE, Lewis EJ, McMenamin RL, Navarro EF, O'Brien MA, O'Reilly M, Reule M, Saxty G, Seavers LCA, Smith D-M, Squires MS, Trewartha G, Walker MT, Woolford AJ-A. 2008. Identification of N-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *J. Med. Chem.* **51**: 4986-4999.
28. Boehm HJ, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, Kostrewa D, Kuehne H, Luebbers T, Meunier-Keller N, Mueller F. 2000. Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J. Med. Chem.* **43**(14): 2664–2674.
29. Choong IC, Lew W, Lee D, Pham P, Burdett MT, Lam JW, Wiesmann C, Luong TN, Fahr B, DeLano WL, McDowell RS, Allen DA, Erlanson DA, Gordon EM, O'Brien T. 2002. Identification of potent and selective small-molecule inhibitors of caspase-3 through the use of extended tethering and structure-based drug design. *J. Med. Chem.* **45**: 5005–5022.
30. O'Brien T, Fahr BT, Sopko MM, Lam JW, Waal ND, Raimundo BC, Purkey HE, Pham P, Romanowski MJ. 2005. Structural analysis of caspase-1 inhibitors derived from tethering. *Acta. Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **61**: 451–458.
31. Forino M, Johnson S, Wong TY, Rozanov DV, Savinov AY, Li W, Fattorusso R, Becattini B, Orry AJ, Jung D, Abagyan RA, Smith JW, Alibek K, Liddington RC, Strongin AY, Pellecchia M. 2005. Efficient synthetic inhibitors of anthrax lethal factor. *Proc. Natl. Acad. Sci. USA* **102**: 9499–9504.
32. Card GL, Blasdel L, England BP, Zhang C, Suzuki Y, Gillette S, Fong D, Ibrahim PN, Artis DR, Bollag G, Milburn MV, Kim S-H, Schlessinger J, Zhang KYJ. 2005. A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. *Nat. Biotechnol.* **23**: 201–207.
33. Apostolakis J, Caflisch A. 1999. Computational ligand design. *Comb. Chem. High Throughput Screen.* **2**: 91–104.
34. Jorgensen WL. 2004. The many roles of computation in drug discovery. *Science* **303**: 1813–1818.
35. Kumar N, Hendriks BS, Janes KA, de Graaf D, Lauffenburger DA. 2006. Applying computational modeling to drug discovery and development. *Drug Discov. Today* **11**(17–18): 806–811.
36. Zoete V, Grosdidier A, Michielin O. 2009. Docking, virtual high throughput screening and in silico fragment-based drug design. *J. Cell Mol. Med.* **13**: 238–248.
37. Momany F, Rone R. 1992. Validation of the general purpose QUANTA 3.2/CHARMm force field. *J. Comput. Chem.* **13**: 888–900.
38. Caflisch A, Fischer S, Karplus M. 1997. Docking by Monte Carlo minimization with a solvation correction: application to an FKBP-substrate complex. *J. Comput. Chem.* **18**: 723–743.
39. Scarsi M, Apostolakis J, Caflisch A. 1997. Continuum electrostatic energies of macromolecules in aqueous solutions. *J. Phys. Chem. A* **101**: 8098–8106.
40. Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A. 1999. Exhaustive docking of molecular fragments on protein binding sites with electrostatic solvation. *Protein. Struct. Funct. Bioinfo.* **37**: 88–105.
41. Majeux N, Scarsi M, Caflisch A. 2001. Efficient electrostatic solvation model for protein-fragment docking. *Protein. Struct. Funct. Bioinfo.* **42**: 256–268.
42. Scarsi M, Majeux N, Caflisch A. 1999. Hydrophobicity at the surface of proteins. *Protein. Struct. Funct. Bioinfo.* **37**: 565–575.
43. Kolb P, Caflisch A. 2006. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J. Med. Chem.* **49**: 7384–7392.
44. Budin N, Majeux N, Caflisch A. 2001. Fragment-based flexible ligand docking by evolutionary opimization. *Biol. Chem.* **382**: 1365–1372.
45. Huang D, Caflisch A. 2004. Efficient evaluation of binding free energy using continuum electrostatic solvation. *J. Med. Chem.* **47**: 5791–5797.
46. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**: 187–217.
47. Brooks BR, Brooks CL III, Mackerell ADJ, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer MS, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. 2009. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**(10): 1545–1614.
48. Cecchini M, Kolb P, Majeux N, Caflisch A. 2004. Automated docking of highly flexible ligands by genetic algorithms: a critical assessment. *J. Comput. Chem.* **25**: 412–422.
49. Kabsch W. 1976. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **A32**: 922–923.
50. Åqvist J, Medina C, Samuelsson J-E. 1994. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **7**: 385–391.

51. Hansson T, Åqvist J. 1995. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* **8**: 1137–1144.
52. Jones-Hertzog DK, Jorgensen WL. 1996. Binding affinities for sulfonamide inhibitors with human thrombin using monte carlo simulations with a linear response method. *J. Med. Chem.* **40**: 1539–1549.
53. Hansson T, Marelius J, Åqvist J. 1998. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.Aided Mol. Design* **12**: 27–35.
54. Wang J, Dixon R, Kollman P. 1999. Ranking ligand binding affinities with avidin: a molecular dynamics-based interaction energy study. *Protein. Struct. Funct. Bioinfo.* **34**: 69–81.
55. Carlson HA, Jorgensen WL. 1995. An extended linear response method for determining free energies of hydration. *J. Phys. Chem.* **99**: 10667–10673.
56. Wall ID, Leach AR, Salt DW, Ford MG, Essex JW. 1999. Binding constants of neuraminidase inhibitors: an investigation of the linear interaction energy method. *J. Med. Chem.* **42**: 5142–5152.
57. Zhou R, Friesner RA, Ghoshs A, Rizzo RC, Jorgensen WJ, Levy RM. 2001. New linear interaction method for binding affinity calculations using a continuum solvent model. *J. Phys. Chem. B* **102**: 10388–10397.
58. Tounge BA, Reynolds CH. 2003. Calculation of the binding affinity of β-secretase inhibitors using the linear interaction energy method. *J. Med. Chem.* **46**: 2074–2082.
59. Tominaga Y, Jorgensen WL. 2004. General model for estimation of the inhibition of protein kinases using Monte Carlo simulations. *J. Med. Chem.* **47**: 2534–2549.
60. Stjernschantz E, Marelius J, Medina C, Jacobsson M, Vermeulen NiPE, Oostenbrink C. 2006. Are automated molecular dynamics simulations and binding free energy calculations realistic tool s in lead optimization? An evaluation of the linear interaction energy (LIE) method. *J. Chem. Inf. Model.* **46**: 1972–1983.
61. Ersmark K, Nervall M, Hamelink E, Janka LK, Clemente JC, Dunn BM, Blackman MJ, Samuelsson B, Åqvist J, Hallberg A. 2005. Synthesis of malarial plasmepsin inhibitors and prediction of binding modes by molecular dynamics simulations. *J. Med. Chem.* **48**: 6090–6106.
62. Warwicker J, Watson HC. 1982. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.* **157**: 671–679.
63. Friedman R, Caflisch A. 2009. Discovery of plasmepsin inhibitors by fragment-based docking and consensus scoring. *Chem. Med. Chem.* **4**: 1317–1326.
64. Ekonomiuk D, Su X-C, Ozawa K, Bodenreider C, Lim SP, Yin Z, Keller TH, Beer D, Patel V, Otting G, Caflisch A, Huang D. 2009. Discovery of a non-peptidic inhibitor of West Nile virus NS3 protease by high-throughput docking. *PloS Negl. Trop. Dis.* **3**: e356.
65. Kolb P, Huang D, Dey F, Caflisch A. 2008. Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J. Med. Chem.* **51**: 1179–1188.
66. Charifson PS, Corkery JJ, Murcko MA, Walters WP. 1999. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **42**: 5100–5109.
67. Klon AE, Glick M, Davies JW. 2004. Combination of a naive Bayes classifier with consensus scoring improves enrichment of high-throughput docking results. *J. Med. Chem.* **47**: 4356–4359.
68. Citron M. 2004. β-Secretase inhibition for the treatment of Alzheimer's disease: promise and challenge. *Trend. Pharmacol. Sci.* **25**: 92–97.
69. Selkoe DJ. 1999. Translating cell biology into therapeutic advances in Alzheimer's disease. *Nature* **399**: A23–A31.
70. Scheff SW, Price DA. 2006. Alzheimer's disease-related alterations in synaptic density: neocortex and hippocampus. *J. Alzheimers Dis.* **9**: 101–115.
71. Lin X, Koelsch G, Wu S, Downs D, Dashti A, Tang J. 2000. Human aspartic protease memapsin 2 cleaves the β-secretase site of β-amyloid precursor protein. *Proc. Natl. Acad. Sci. USA* **97**: 1456–1460.
72. Petkova AT, Leapman RD, Guo Z, Yau WM, Mattson MP, Tycko R. 2005. Self-propagating, molecular-level polymorphism in Alzheimer's β-amyloid fibrils. *Science* **307**: 262–265.
73. Cumming JN, Iserloh U, Kennedy ME. 2004. Design and development of BACE-1 inhibitors. *Curr. Opin. Drug. Discov. Devel.* **7**: 536–556.
74. Stanton MG, Stauffer SR, Gregro AR, Steinbeiser M, Nantermet P, Sankaranarayanan S, Price EA, Wu G, Crouthamel M-C, Ellis J, Lai M-T, Espeseth AS, Shi X-P, Jin L, Colussi D, Pietrak B, Huang Q, Xu M, Simon AJ, Graham SL, Vacca JP, Selnick H. 2007. Discovery of isonicotinamide derived beta-secretase inhibitors: in vivo reduction of beta-amyloid. *J. Med. Chem.* **50**: 3431–3433.
75. Hussain I, Hawkins J, Harrison D, Hille C, Wayne G, Cutler L, Buck T, Walter D, Demont E, Howes C, Naylor A, Jeffrey P, Gonzalez MI, Dingwall C, Michel A, Redshaw S, Davis JB. 2007. Oral administration of a potent and selective non-peptidic BACE-1 inhibitor decreases beta-cleavage of amyloid precursor protein and amyloid-beta production in vivo. *J. Neurochem.* **100**: 802–809.
76. Sankaranarayanan S, Price EA, Wu G, Crouthamel M-C, Shi X-P, Tugusheva K, Tyler KX, Kahana J, Ellis J, Jin L, Steele T, Stachel S, Coburn C, Simon AJ. 2008. In vivo beta-secretase 1 inhibition leads to brain Abeta lowering and increased alpha-secretase processing of amyloid precursor protein without effect on neuregulin-1. *J. Pharmacol. Exp. Ther.* **324**: 957–969.
77. Sankaranarayanan S, Holahan MA, Colussi D, Crouthamel M-C, Devanarayan V, Ellis J, Espeseth A, Gates AT, Graham SL, Gregro AR, Hazuda D, Hochman JH, Holloway K, Jin L, Kahana J, tain Lai M, Lineberger J, McGaughey G, Moore KP, Nantermet P, Pietrak B, Price EA, Rajapakse H, Stauffer S, Steinbeiser MA, Seabrook G, Selnick HG, Shi X-P, Stanton MG, Swestock J, Tugusheva K, Tyler KX, Vacca JP, Wong J, Wu G, Xu M, Cook JJ, Simon AJ. 2009. First demonstration of cerebrospinal fluid and plasma A beta lowering with oral administration of a beta-site amyloid precursor protein-cleaving enzyme 1 inhibitor in nonhuman primates. *J. Pharmacol. Exp. Ther.* **328**: 131–140.
78. Asai M, Hattori C, Iwata N, Saido TC, Sasagawa N, Szabo B, Hashimoto Y, Maruyama K, ichi Tanuma S, Kiso Y, Ishiura S. 2006. The novel beta-secretase inhibitor KMI-429 reduces amyloid beta peptide production in amyloid precursor protein transgenic and wild-type mice. *J. Neurochem.* **96**: 533–540.
79. Roggo S. 2002. Inhibition of BACE, a promising approach to Alzheimer's disease therapy. *Curr. Top. Med. Chem.* **2**: 359–370.
80. Middendorp O, Lüthi U, Hausch F, Barberis A. 2004. Searching for the most effective screening system to identify cell-active inhibitors of β-secretase. *Biol. Chem.* **385**: 481–485.
81. Ghosh AK, Bilcer G, Hong L, Koelsch G, Tang J. 2007. Memapsin 2 (beta-secretase) inhibitor drug, between fantasy and reality. *Curr. Alzheimer Res.* **4**: 418–422.
82. Coburn CA, Stachel SJ, Li YM, Rush DM, Steele TG, Chen-Dodson E, Holloway MKe. 2004. Identification of a small molecule nonpeptide active site β-secretase inhibitor that displays a nontraditional binding mode for aspartyl proteases. *J. Med. Chem.* **47**: 6117–6119.
83. Huang D, Lüthi U, Kolb P, Edler K, Cecchini M, Audetat S, Barberis A, Caflisch A. 2005. Discovery of cell-permeable non-peptide inhibitors of β-secretase by high-throughput docking and continuum electrostatics calculations. *J. Med. Chem.* **48**: 5108–5111.
84. Huang D, Lüthi U, Kolb P, Cecchini M, Barberis A, Caflisch A. 2006. In silico discovery of β-secretase inhibitors. *J. Am. Chem. Soc.* **128**: 5436–5443.
85. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI. 2005. The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature* **434**(7030): 214–217.
86. Egan TJ, Kaschula CH. 2007. Strategies to reverse drug resistance in malaria. *Curr. Opin. Infect. Dis.* **20**(6): 598–604.
87. Rosenthal PJ. 1998. Proteases of malaria parasites: new targets for chemotherapy. *Emerg. Infect. Dis.* **4**: 49–57.
88. Ersmark K, Samuelsson B, Hallberg A. 2006. Plasmepsins as potential targets for new antimalarial therapy. *Med. Res. Rev.* **26**: 626–666.
89. Hammer SM, Eron JJJ, Reiss P, Schooley RT, Thompson MA, Walmsley S, Cahn P, Fischl MA, Gatell JM, Hirsch MS, Jacobsen DM, Montaner JSG, Richman DD, Yeni PG, Volberding PA. 2008. Antiretroviral treatment of adult HIV infection: 2008 recommendations of the International AIDS Society-USA panel. *JAMA* **300**: 555–570.
90. Gaddam KK, Oparil S. 2008. Renin inhibition: should it supplant ACE inhibitors and ARBS in high risk patients? *Curr. Opin. Nephrol. Hypertens* **17**: 484–490.
91. Division of Vector-Borne Infectious Disease. West Nile virus homepage. Centers for Disease Control and Prevention http://www.cdc.gov/ncidod/dvbid/westnile 2007.
92. Mukhopadhyay S, Kuhn RJ, Rossmann MG. 2005. A structural perspective of the flavivirus life cycle. *Nat. Rev. Microbiol.* **3**: 13–22.

93. Chappell KJ, Stoermer MJ, Fairlie DP, Young PR. 2006. Insights to substrate binding and processing by West Nile Virus NS3 protease through combined modeling, protease mutagenesis, and kinetic studies. *J. Biol. Chem.* **281**: 38448–38458.
94. Perni RB, Almquist SJ, Byrn RA, Chandorkar G, Chaturvedi PR, Courtney LF, Decker CJ, Dinehart K, Gates CA, Harbeson SL, Heiser A, Kalkeri G, Kolaczkowski E, Lin K, Luong Y-P, Rao BG, Taylor WP, Thomson JA, Tung RD, Wei Y, Kwong AD, Lin C. 2006. Preclinical profile of VX-950, a potent, selective, and orally bioavailable inhibitor of hepatitis C virus NS3-4A serine protease. *Antimicrob. Agents Chemother.* **50**: 899–909.
95. Tomei L, Failla C, Santolini E, Francesco RD, Monica NL. 1993. NS3 is a serine protease required for processing of hepatitis C virus polyprotein. *J. Virol.* **67**: 4017–4026.
96. Malcolm BA, Liu R, Lahser F, Agrawal S, Belanger B, Butkiewicz N, Chase R, Gheyas F, Hart A, Hesk D, Ingravallo P, Jiang C, Kong R, Lu J, Pichardo J, Prongay A, Skelton A, Tong X, Venkatraman S, Xia E, Girijavallabhan V, Njoroge FG. 2006. SCH 503034, a mechanism-based inhibitor of hepatitis C virus NS3 protease, suppresses polyprotein maturation and enhances the antiviral activity of alpha interferon in replicon cells. *Antimicrob. Agents Chemother.* **50**: 1013–1020.
97. Seiwert SD, Andrews SW, Jiang Y, Serebryany V, Tan H, Kossen K, Rajagopalan PTR, Misialek S, Stevens SK, Stoycheva A, Hong J, Lim SR, Qin X, Rieger R, Condroski KR, Zhang H, Do MG, Lemieux C, Hingorani GP, Hartley DP, Josey JA, Pan L, Beigelman L, Blatt LM. 2008. Preclinical characteristics of the hepatitis C virus NS3/4A protease inhibitor itmn-191 (r7227). *Antimicrob. Agents Chemother.* **52**(12): 4432–4441.
98. Ekonomiuk D, Su X-C, Bodenreider C, Lim SP, Otting G, Huang D, Caflisch A. 2009. Flaviviral protease inhibitors identified by fragment-based library docking into a structure generated by molecular dynamics. *J. Med. Chem.* **52**: 4860–4868. DOI: 10.1021/jm900448m
99. Hopkins AL, Groom CR, Alex A. 2004. Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **9**(10): 430–4431.
100. Bodenreider C, Beer D, Keller T, Sonntag S, Wen D, Yap L, Hoe Yau Y, Geifman Shochat S, Huang D, Zhou T, Caflisch A, Sux, Ozawa K, Otting G, Vasudevan S, Lescar J, Lim S. 2009. Identification of lead-like and non-lead-like inhibitors for the dengue virus NS2B/NS3 protease by tryptophan fluorescence. *Analytical Biochemistry*, DOI: 10.1016/j.ab.2009.08.013.
101. Mort JS, Cathepsin B. 2004. Handbook of Proteolytic Enzymes, (2nd ed.) Barrett AJ, Rawlings ND, Woessner JF Jr (eds). Elsevier: London, UK.
102. Aronson NNJ, Barrett AJ. 1978. The specificity of cathepsin B. Hydrolysis of glucagon at the C-terminus by a peptidyldipeptidase mechanism. *Biochem. J.* **171**(3): 759–765.
103. Bond JS, Barrett AJ. 1980. Degradation of fructose-1,6-bisphosphate aldolase by cathepsin B. *Biochem. J.* **189**(1): 17–25.
104. Takahashi T, Dehdarani AH, Yonezawa S, Tang J. 1986. Porcine spleen cathepsin B is an exopeptidase. *J. Biol. Chem.* **261**(20): 9375–9381.
105. Rowan AD, Feng R, Konishi Y, Mort JS. 1993. Demonstration by electrospray mass spectrometry that the peptidyldipeptidase activity of cathepsin B is capable of rat cathepsin B C-terminal processing. *Biochem. J.* **294** (Pt 3)(NIL): 923–927.
106. Kukor Z, Mayerle J, Kruger B, Toth M, Steed PM, Halangk W, Lerch MM, Sahin-Toth M. 2002. Presence of cathepsin B in the human pancreatic secretory pathway and its role in trypsinogen activation during hereditary pancreatitis. *J. Biol. Chem.* **277**: 21389–21396.
107. Broker LE, Kruyt FAE, Giaccone G. 2005. Cell death independent of caspases: a review. *Clin. Cancer Res.* **11**: 3155–3162.
108. Yan S, Sloane BF. 2003. Molecular regulation of human cathepsin B: implication in pathologies. *Biol. Chem.* **384**: 845–854.
109. Mohamed MM, Sloane BF. 2006. Cysteine cathepsins: multifunctional enzymes in cancer. *Nat. Rev. Cancer* **6**: 764–775.
110. Lenarcic B, Gabrijelcic D, Rozman B, Drobnic-Kosorok M, Turk V. 1988. Human cathepsin B and cysteine proteinase inhibitors (CPIs) in inflammatory and metabolic joint diseases. *Biol. Chem. Hoppe. Seyler* **369**(Suppl): 257–261.
111. Baici A, Lang A, Horler D, Kissling R, Merlin C. 1995. Cathepsin B in osteoarthritis: cytochemical and histochemical analysis of human femoral head cartilage. *Ann. Rheum. Dis.* **54**: 289–297.
112. Baici A, Horler D, Lang A, Merlin C, Kissling R. 1995. Cathepsin B in osteoarthritis: zonal variation of enzyme activity in human femoral head cartilage. *Ann. Rheum. Dis.* **54**: 281–288.
113. Schenker P, Alfarano P, Kolb P, Caflisch A, Baici A. 2008. A double-headed cathepsin B inhibitor devoid of warhead. *Protein Sci.* **17**(12): 2145–2155.
114. Cheng N, Brantley DM, Chen J. 2002. The ephrins and Eph receptors in angiogenesis. *Cytokine Growth Factor Rev.* **13**(1): 75–85.
115. Kolb P, Berset C, Huang D, Caflisch A. 2008. Structure-based tailoring of compound libraries for high-throughput screening: Discovery of novel EphB4 kinase inhibitors. *Protein. Struct. Funct. Bioinfo.* **73**: 11–18.
116. Lafleur K, Huang D, Zhou T, Caflisch A, Nevado C. 2009. Structure-based optimization of potent and selective inhibitors of the tyrosine kinase Ephb4. *J. Med. Chem.* (in press).
117. Sherr CJ. 1996. Cancer cell cycles. *Science* **274**: 1672–11677.
118. Nurse P, Masui Y, Hartwell L. 1998. Understanding the cell cycle. *Nat. Med.* **4**: 1103–1106.
119. Morgan DO. 1997. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu. Rev. Cell Dev. Biol.* **13**: 261–291.
120. Davis ST, Benson BG, Bramson HN, Chapman DE, Dickerson SH, Dold KM, Eberwein DJ, Edelstein M, Frye SV Jr, Griffin RTG, Harris RJ, Hassell PA, Holmes AM, Hunter WD, Knick RN, Lackey VB, Lovejoy K, Luzzio B, Murray MJ, Parker D, Rocque P, Shewchuk WJ, Veal L, Walker JM, Walker DH, Kuyper LF. 2001. Prevention of chemotherapy-induced alopecia in rats by CDK inhibitors. *Science* **291**: 134–137.
121. Fink T, Bruggesser H, Reymond J-L. 2005. Virtual exploration of the small-molecule chemical universe below 160 Daltons. *Angew. Chem. Int. Ed. Engl.* **44**(10): 1504–1508.
122. Bohacek RS, McMartin C, Guida WC. 1996. The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* **16**(1): 3–50.
123. Kolb P, Rosenbaum D, Irwin J, Fung JJ, Kobilka BK, Shoichet BK. 2009. Structure-based discovery of β_2-adrenergic receptor ligands. *Proc. Natl. Acad. Sci. USA* **106**: 6843–6848.
124. Vangrevelinghe E, Zimmermann K, Schoepfer J, Portmann R, Fabbro D, Furet P. 2003. Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *J. Med. Chem.* **46**(13): 2656–2662.
125. Zhou T, Huang D, Caflisch A. 2008. Is quantum mechanics necessary for predicting binding free energy? *J. Med. Chem.* **51**: 4280–4288.
126. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**: 10037–10041.

