

# Structure-based tailoring of compound libraries for high-throughput screening: Discovery of novel EphB4 kinase inhibitors

Peter Kolb,<sup>1</sup> Catherine Berset Kipouros,<sup>2</sup> Danzhi Huang,<sup>1</sup> and Amedeo Caflisch<sup>1\*</sup>

<sup>1</sup>Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland

<sup>2</sup>Oncalis AG, Biotech Center Zurich, CH-8952 Schlieren, Switzerland

## ABSTRACT

High-throughput docking is a computational tool frequently used to discover small-molecule inhibitors of enzymes or receptors of known three-dimensional structure. Because of the large number of molecules in chemical libraries, automatic procedures to prune multimillion compound collections are useful for high-throughput docking and necessary for *in vitro* screening. Here, we propose an anchor-based library tailoring approach (termed ALTA) to focus a chemical library by docking and prioritizing molecular fragments according to their binding energy which includes continuum electrostatics solvation. In principle, ALTA does not require prior knowledge of known inhibitors, but receptor-based pharmacophore information (hydrogen bonds with the hinge region) is additionally used here to identify molecules with optimal anchor fragments for the ATP-binding site of the EphB4 receptor tyrosine kinase. The 21,418 molecules of the focused library (from an initial collection of about 730,000) are docked into EphB4 and ranked by force-field-based energy including electrostatic solvation. Among the 43 compounds tested *in vitro*, eight molecules originating from two different anchors show low-micromolar activity in a fluorescence-based enzymatic assay. Four of them are active in a cell-based assay and are potential anti-angiogenic compounds.

Proteins 2008; 73:11–18.  
© 2008 Wiley-Liss, Inc.

**Key words:** EphB4; tyrosine kinases; angiogenesis; cancer; fragment-based docking; high-throughput screening.

## INTRODUCTION

Docking is a successful method in computer-aided drug design.<sup>1–5</sup> However, there is a large and increasing discrepancy between the number of small molecules available in computer-readable format (on the order of  $10^7$ – $10^8$  compounds,<sup>6</sup> not counting those generated by *de novo* design programs) and the amount of molecules that can be efficiently processed by high-throughput approaches *in silico* or *in vitro* (on the order of  $10^5$ – $10^6$ ). Moreover, it is highly inefficient to use computational resources or robotic systems and chemical reagents for compounds that have a low chance of binding to the target protein. The essential question is how to utilize (structural) information of the target to preselect the molecules that are most likely to show binding and inhibitory activity. One efficient approach exploits the available experimental knowledge of known inhibitors by using pharmacophore constraints to preselect compounds for docking, thereby reducing computation times. In this way, inhibitors of aldose reductase<sup>7</sup> and the MDM2 oncoprotein<sup>8</sup> were identified.

Here, we introduce an efficient computational method to focus a library (termed ALTA for anchor-based library tailoring) and present an application which has resulted in the discovery of inhibitors of the erythropoietin producing human hepatocellular carcinoma receptor tyrosine kinase B4 (EphB4). The basic strategy of the ALTA approach (Fig. 1) has been inspired by *in silico* fragment-based approaches<sup>10,13</sup> and *in vitro* “needle screening” procedures.<sup>14</sup> Needle- or anchor-based *in vitro* screening has been applied to a wide range of enzymes including thrombin,<sup>14</sup> DNA gyrase,<sup>15</sup> and protein tyrosine kinases.<sup>16,17</sup> Therefore, although only one kinase is used in this study to validate the approach, the applicability of ALTA extends to any target enzyme for which the three-dimensional structure and substrate binding site are known.

The receptor tyrosine kinase EphB4 is a highly attractive angiogenic target involved in many types of cancer.<sup>18</sup> It seems to be rather recalcitrant to inhibition because, despite its potential therapeutic importance, only two

Additional Supporting Information may be found in the online version of this article.

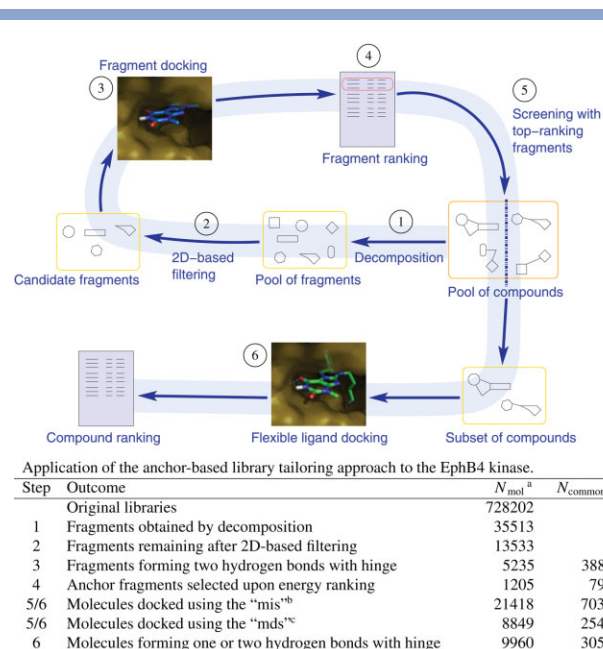
Peter Kolb's current address is Department of Pharmaceutical Chemistry, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94158, USA.

Grant sponsor: Swiss Commission for Technology and Innovation (KTI/CTI).

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

Received 25 October 2007; Revised 29 January 2008; Accepted 5 February 2008

Published online 2 April 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22028

**Figure 1**

Graphical representation of the workflow of the ALTA procedure (top) and its application to EphB4 (bottom). The first step is the automatic decomposition<sup>9</sup> of a library of compounds (orange rectangle) to obtain the pool of fragments. Afterward, fragments selected based on the binding site features are docked<sup>10</sup> and ranked according to their binding energy.<sup>11</sup> Poses for molecules that contain at least one of the top-ranking fragments are then generated by flexible-ligand docking.<sup>12</sup> <sup>a</sup>Number of fragments/compounds processed in the individual steps. Docking (Steps 3 and 6) was carried out in parallel on two structures of EphB4 differing only in the orientation of the hydroxyl group of Thr693 in the ATP-binding site. The value of  $N_{mol}$  is the number of unique fragments (in Steps 3–5) or unique molecules (in Step 6) originating from the docking into the two structures. In contrast,  $N_{common}$  is the number of fragments or molecules that were docked to both structures. As an example, there are 1205 unique anchor fragments, 795 of which have favorable binding energy in both EphB4 structures, upon merging the two sets of 1000 fragments (one set for each Thr693 orientation) with most favorable binding energy calculated by SEED. <sup>b</sup>Most interesting set (mis): flexible-ligand docking using the three fragments with the highest chemical richness<sup>9</sup> as anchors. <sup>c</sup>Maximum diversity set (mds): flexible-ligand docking using the three fragments which are most dissimilar to each other as anchors. The compounds docked using the mds are a subset of the compounds docked using the mis.

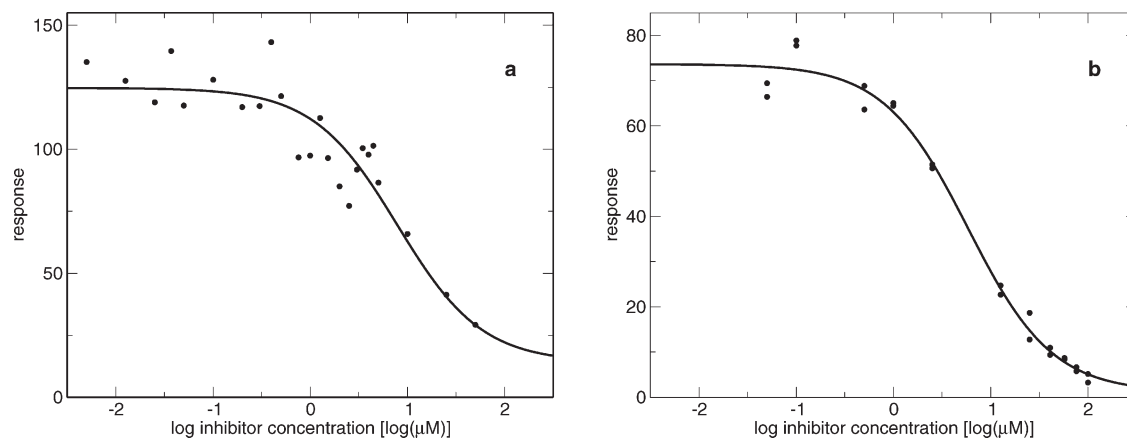
series of nonpeptidic small molecule inhibitors have been reported in the literature up to date.<sup>19</sup> This work reports on the efficient docking of a focused compound library, tailored by ALTA to the ATP-binding site of EphB4. The docking of the focused library results in the identification of several small molecules that are promising anti-angiogenic lead compounds against EphB4.

## METHODS

### Computational methods

The procedure for ALTA starts from a large pool of compounds and consists of six steps, which are illustrated in Figure 1 together with the statistics of the application

to EphB4. It is important to underline that although a suite of in-house developed softwares was used in this study, the basic workflow can be reproduced with any combination of softwares that is capable of carrying out all the steps. Three libraries were combined for a total of 728,202 compounds: the NCI database,<sup>20</sup> the ChemDiv library (Moscow, Russia), and the synthetic and natural compound library of InterBioScreen (Moscow, Russia). In Step 1 of ALTA, the compounds were decomposed into predominantly rigid fragments by automatic identification and cutting of rotatable bonds.<sup>9</sup> The three fragments with the highest chemical richness (sum over the entries in the molecular fingerprint; cf. Fig. 2 in Ref. 9) were selected for every compound. From the resulting set of 35,513 unique fragments, candidate anchors were chosen which fulfil the constraints imposed by the binding site (Step 2 in Fig. 1). Although it is not a requirement of ALTA to take pharmacophore constraints into account, doing so will help to discard at an early stage fragments which are unlikely to bind. In this study, the ability of a fragment to form the characteristic bidentate hydrogen bonds of ATP with the hinge region of a kinase<sup>21</sup> was used. Fragments that contained at least one ring, one hydrogen bond donor (nitrogen, oxygen, or sulfur connected to a hydrogen), and one acceptor (nitrogen with a lone pair or oxygen) were docked with a rigid-fragment docking program<sup>10,11</sup> (Step 3 in Fig. 1). The presence of a ring was required to select only fragments with a certain rigidity and thus fixed relative orientation of the donor and acceptor group(s). Docked fragments were considered to form the bidentate hydrogen bonds if a matching donor and acceptor were found within a radius of 3 Å around the carbonyl oxygen of Glu694 and the amide nitrogen of Met696, respectively. Glu694 and Met696 are the first and last residue in the hinge region of EphB4. These fragments were ranked according to their binding energy (Step 4 in Fig. 1), which is the sum of van der Waals interaction, electrostatic desolvation<sup>22</sup> of both the receptor and the fragment, and the screened electrostatic interaction.<sup>10</sup> There are no terms reflecting the internal energy, since fragments are treated as rigid. Using the 1205 top-ranking fragments as queries, 21,418 compounds were retrieved from the composite library of 728,202 molecules (Step 5 in Fig. 1). This search used an algorithm for the identification of a graph-subgraph isomorphism<sup>22</sup> which was implemented in DAIM,<sup>9</sup> adapting the VFLib Graph Matching Library, version 2.0 in Ref. 23. Finally, in Step 6, a suite of programs was used for fragment-based<sup>9,10</sup> flexible-ligand docking by genetic algorithm optimization.<sup>12</sup> The protein was kept rigid during docking, but two structures with different orientations of the hydroxyl group of Thr693 were used for both fragment and compound docking in order to allow for a greater variety of eventual hydrogen bond interactions with potential ligands. The third column in the table of Figure 1 lists the numbers of unique fragments (or compounds), whereas the fourth



**Figure 2**

Plot of the response [arbitrary units] versus the concentration of compound 2 on a logarithmic scale. Data obtained with the Omnia Tyr Recombinant Kit KNZ4051 (Biosource, USA) without (a) and with (b) the addition of detergent (0.01% of Triton X-100). Values in panel (b) were measured in duplicates. The concentration of ATP in both assays was 125  $\mu$ M, which is close to its  $K_m$ .

column lists the fragments selected for (or compounds docked into) both structures.

### EphB4 model

Since the crystal structure of the kinase domain of EphB4 is not known, a homology model was built using the structure of EphB2 (mouse, PDB entry 1JPA<sup>24</sup>) as template. The overall amino acid sequence identity between the human EphB4 sequence obtained from SWISS-PROT (accession code P54760) and the sequence derived from the mouse EphB2 structure is 88.4% and there are no gaps or insertions in the aligned region. An additionally generated binary sequence alignment with the sequence derived from the human Eph kinase structure EphA2 (PDB entry 1MQB) revealed a lower sequence identity of 63.1% and also three short regions containing gaps or insertions. Therefore, only the sequence alignment between EphB4 (human) and EphB2 (mouse) was used in the initial phase of the homology modeling procedure. All sequence alignments were performed using the program ClustalW.<sup>25</sup> The 1JPA crystal structure comprises two chemically identical subunits in the crystallographic asymmetric unit. A structural superposition of the two subunits results in an average root-mean square deviation of 0.32 Å for 269  $C_\alpha$  atoms. Since subunit A of the 1JPA crystal structure has better main-chain dihedral angles ( $\phi$  and  $\psi$ ) and lower B factors, this subunit was initially chosen as the template structure. However, the two subunits of the 1JPA structure show some clear structural differences around the active site region. The differing amino acid residues were analyzed with respect to B factors, possible contacts, and stereochemical criteria. Subsequently, the side-chain rotamer

conformations of the initial template structure (subunit A) were replaced by the corresponding side-chain conformations of subunit B of the 1JPA crystal structure, if the latter were better defined according to the aforementioned criteria. The resulting modified structure was then used as template for homology modeling and a total of 100 different models were generated using the program Modeller.<sup>26,27</sup>

The obtained initial models were ranked and analyzed based on internal energy and stereochemical quality. The best model was manually modified by comparison with the active sites of the two known Eph kinase structures (PDB entry codes 1JPA and 1MQB) and other related kinase structures (PDB entry codes 1BYG, 1FMK, 1FPU, 1IEP, 1M14, 1M17, 1M52, 1MP8, 1OPK, 1OPL, and 2SRC) found by PSI-BLAST<sup>28</sup> and DALI<sup>29</sup> searches of protein structure databases. The conformational information contained in these homologous structures was used for manually adjusting side-chain conformations of conserved or similar amino acid residues. If feasible on the basis of energetic and packing criteria, side-chain  $\chi$ -angles of the respective amino acid residues were completely or partially conserved. Or else, statistically preferred  $\chi$ -angles were chosen and favorable hydrophobic or polar contacts were taken into consideration. After each manual change, the side-chain in question was minimized with CHARMM<sup>30</sup> and the CHARMM22<sup>31</sup> force field. Moreover, an ATP molecule was modeled into the active site to avoid structural changes during minimization. Hydrogen atoms (considering appropriate ionization states for acidic and basic amino acid residues), CHARMM22<sup>31</sup> atom types, and partial charges were assigned to the protein and ATP using the program WITNOTP and the MPEOE method.<sup>32,33</sup> This rebuilding and refinement with

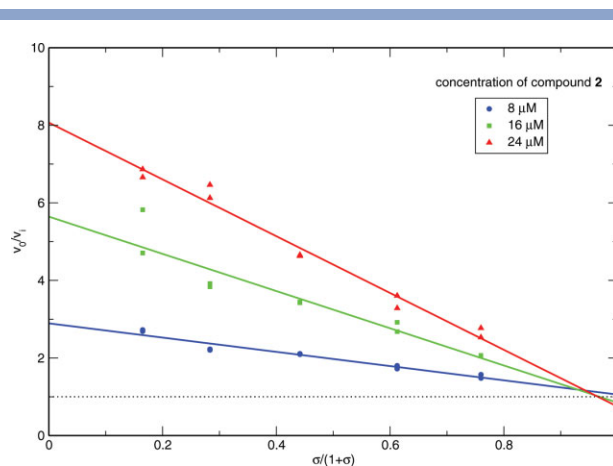
CHARMM<sup>30</sup> using the CHARMM22<sup>31</sup> force field led to a model exhibiting excellent stereochemical quality, with 92.9% of the  $\phi/\psi$  values in the most favored regions and 6.3% in additionally allowed regions of the Ramachandran plot as evaluated with the program PROCHECK.<sup>34</sup> Finally, the ATP molecule was removed.

### Minimization and ranking

After docking, multiple poses of each compound were minimized with CHARMM<sup>30</sup> and the CHARMM22<sup>31</sup> force field. Minimized poses were filtered according to thresholds in van der Waals energy and van der Waals efficiency (the ratio of the van der Waals energy and the molecular weight) (Supplementary Materials). A general linear interaction energy with continuum electrostatics (LIECE<sup>35</sup>) model<sup>36</sup> was used for the final ranking (see the next paragraph).

### LIECE model for kinases

The two-parameter LIECE model  $\Delta G = \alpha \Delta E_{\text{vdW}} + \beta \Delta G_{\text{elec}}$  was developed using a composite set of 165 known inhibitors, that is, 73, 51, and 41 inhibitors of cyclin-dependent kinase 2 (CDK2), lymphocyte-specific kinase (Lck), and p38 mitogen-activated protein kinase (p38), respectively. It showed a very low root-mean-square of the error (rmse) of only 1.03 kcal/mol for the 165 inhibitors used as training set in the fitting and a leave-one-out cross-validated  $q^2$  of 0.74.<sup>36</sup> The values of  $\alpha$  and  $\beta$  for this kinase LIECE model are  $0.2898 \pm 0.0075$  (model with all 165 inhibitors  $\pm$  standard deviation over 165 LIECE models obtained by leave-one-out) and  $0.0442 \pm 0.0074$ , respectively. The values of these parameters are similar to those obtained for the three models derived by fitting to each of the three kinases individually. The two-parameter three-protein LIECE model was also cross validated on a set of compounds not used for fitting, namely 128 known inhibitors of the epidermal growth factor receptor (EGFR) tyrosine kinase.<sup>37</sup> On this set, the model yielded an rmse of 1.46 kcal/mol. This transferability of parameters is in contrast to our previous study,<sup>35</sup> where parameters were not transferable between HIV-1 protease ( $\alpha = 0.1690$  and  $\beta = 0.0168$  using 24 peptidic inhibitors) and  $\beta$ -secretase ( $\alpha = 0.2737$  and  $\beta = 0.1795$  using 13 peptidic inhibitors). The nontransferability between retroviral and mammalian aspartic proteases is due to the fact that they differ not only in overall shape (HIV-1 protease is a homodimer of  $99 \times 2$  residues while human  $\beta$ -secretase is a monomer of 501 residues) but also, and more importantly for LIECE, in the substrate-binding site. This difference stems from the different polypeptide sequences of their substrates. In contrast, the ATP-binding site of kinases, which was targeted in this study, is highly conserved in structure.<sup>21</sup>



**Figure 3**

Specific velocity plot<sup>38</sup> for compound 2.  $v_0/v_i$  is the ratio of the initial velocities for the noninhibited and inhibited reactions at a given ATP concentration. All values were measured in duplicates, but some of the pairs of data points (especially the blue ones) almost completely overlap on the plot.  $\sigma$  is the ratio of the ATP concentration and its  $K_m$ , which was 158  $\mu\text{M}$  in this experiment. ATP concentrations ranged from 31.25 to 500  $\mu\text{M}$ . Intersecting of all curves close to (1,1) indicates that compound 2 is ATP-competitive. Measurements were carried out by the Omnia kinetic assay as described.

### Kinetic assay

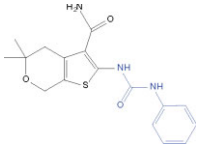
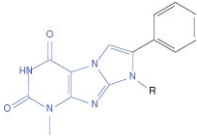
Compound 2 was tested in the Omnia Tyr Recombinant Kit KNZ4051 (Biosource, USA) at different concentrations between 50 nM and 100  $\mu\text{M}$  in a Corning 96-well microtiter plate. Fluorescence progress curves were measured upon excitation at 360 nm and emission at 485 nm. The assay contained a final concentration of EphB4 and ATP of 25 ng/ $\mu\text{L}$  and 125  $\mu\text{M}$  (which is near its  $K_m$ ), respectively, and was run at 30°C for 1 h. To provide evidence for specific binding, a second assay was performed in which 0.01% of Triton X-100 was added to the reaction mixture (Fig. 2). Moreover, to show the competitive behaviour of compound 2 toward ATP, different ATP concentrations (ranging from 31.25 to 500  $\mu\text{M}$ ) were used (Fig. 3). IC<sub>50</sub> values (inhibitor concentration at which enzyme activity is reduced by 50%) and fitted curves were determined with GraphPad Prism 5.0 (GraphPad Software).

### Phospho-EphB4 ELISA of CHO cells

The ELISA assay was performed essentially as described in Ref. 39. Briefly, Chinese hamster ovary (CHO)-FRT cells were stably transfected with myc-tagged full-length human EphB4 according to the Flip-In System protocol (Invitrogen, USA). A clone that expressed myc-tagged human EphB4 and showed inducible autophosphorylation of EphB4 upon stimulation with mouse ephrinB2-Fc (mouse ephrinB2 fused to the Fc region of human IgG) was selected. Cells of this clone were preincubated with



**Table 1**  
Experimental Validation of EphB4 Inhibitors Identified by ALTA

Compound	Structure	MW (Da)	Enzymatic assays			Cell-based assay	Predicted $\Delta G^{\circ}$ (kcal/mol)
			IC <sub>50</sub> <sup>a</sup> ( $\mu M$ )	IC <sub>50</sub> <sup>b</sup> ( $\mu M$ )	IC <sub>50</sub> <sup>c</sup> ( $\mu M$ )	CHO Ref. 39 (%) <sup>d</sup>	
1		349	76	n.d.	n.d.	n.d.	-8.8
2-8							
	R=						
2	-(CH <sub>2</sub> ) <sub>4</sub> OH	353	1.4, 1.8	6.8	5.6, 7.9	neg.	-11.0
3	-(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	337	1.5, 1.5	8.3	n.d.	neg.	-10.7
4	-(CH <sub>2</sub> ) <sub>2</sub> CH <sub>3</sub>	323	1.9	n.d.	n.d.	neg.	
5	<i>ortho</i> -methoxy-phenyl	387	27	n.d.	n.d.	14% at 20 $\mu M$	
6	-(CH <sub>2</sub> ) <sub>3</sub> OCH <sub>3</sub>	353	50	n.d.	n.d.	28% at 20 $\mu M$	
7	-(CH <sub>2</sub> ) <sub>2</sub> Ph	385	30% at 14 $\mu M$ <sup>d</sup>	n.d.	n.d.	34% at 20 $\mu M$	
8	-(CH <sub>2</sub> ) <sub>2</sub> - <i>meta</i> -methylphenyl	399	29% at 5 $\mu M$ <sup>d</sup>	n.d.	n.d.	14% at 20 $\mu M$	

The fragments used to retrieve the molecules are emphasized in blue.

<sup>a</sup>Z'Lyte FRET-based enzymatic assay. Compounds 2 and 3 were tested twice.

<sup>b</sup>Cerep FRET-based enzymatic assay (all assays are described in the Suppl. Mat.).

<sup>c</sup>Omnia kinetic assay.

<sup>d</sup>Percent inhibition at specified concentration.

<sup>e</sup>Free energy of binding calculated by LIECE (see Methods section).

EphB4 inhibitors for 15 min and then stimulated with preclustered ephrinB2-Fc for 45 min at 37°C. Cells were lysed, transferred to a 96-well plate coated with anti-myc antibody and incubated overnight at 4°C. After incubation with antiphosphotyrosine antibody, reactions were developed with BM Blue peroxidase substrate, and autophosphorylation of EphB4 was quantified by measuring absorption at 450 nm.

## RESULTS AND DISCUSSION

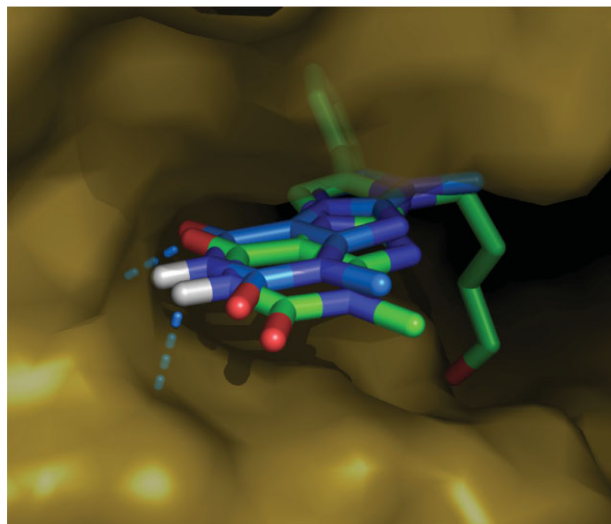
### Library tailoring by ALTA using the EphB4 structure

Only fragments involved in two or more hydrogen bonds with the EphB4 hinge region were considered in ALTA. Fragments were ranked according to binding energy calculated by SEED (which includes continuum electrostatics solvation). The union of the two sets of 1000 fragments (one set for each protein structure differing in the orientation of the hydroxyl group of Thr693, see Methods section) with the most favorable calculated binding energy consists of 1205 unique fragments. A subset of 21,418 compounds was retrieved from the composite library of 728,202 molecules (Step 5 in Fig. 1) using these 1205 fragments.

### In vitro validation

Forty compounds from the high-throughput docking of the focused library of 21,418 molecules were selected upon ranking (Step 6 in Fig. 1) and tested first in a fluorescence-resonance energy transfer (FRET)-based enzymatic assay (see Supplementary Materials). Ten of these could not be measured, because they interfered with the fluorescence read-out. A compound with a phenylurea anchor (**1** in Table 1) showed an IC<sub>50</sub> of 76  $\mu M$  in the FRET-based enzymatic assay. In the same assay, another compound with a three-ring system as anchor (**2** in Table 1) showed an IC<sub>50</sub> of about 1.6  $\mu M$ , with a molecular weight of only 353 Da. Five more compounds with different anchors inhibited the activity of EphB4 by 15–40% at a concentration of 125  $\mu M$  (data not shown).

The ligand efficiency (LE =  $-\Delta G_{\text{binding}}^{\text{exp}}/\text{HAC}$ , where HAC is the number of heavy atoms) is a useful measure for lead selection.<sup>40</sup> Notably, inhibitor **2** has an LE value of 0.3 kcal/mol per heavy atom suggesting that it is an interesting compound for further development. To evaluate its cell permeability and cellular activity, compound **2** was tested in CHO cells for inhibition of EphB4 autophosphorylation in a mammalian cell-based environment. It showed only mild inhibitory effects in CHO cells at a concentration of 20  $\mu M$ . To further investigate cellular activity and determine whether lack thereof is a



**Figure 4**

Predicted binding mode of compound **2** (carbon atoms in green) and its anchor fragment (carbon atoms in light blue) in a homology model of EphB4 (gold surface). The hydrogen bonds with the hinge region are shown in light blue dashes. Note the significant overlap in the binding mode of compound and fragment. Figure prepared with PyMOL (DeLano Scientific, USA).

general problem of this class of molecules, 13 molecules containing the scaffold of compound **2** were purchased (of these, only compound **3** was present in the original library. It had been ranked among the first 300 molecules but was not purchased in the first round because it was considered too similar to compound **2**). Two of the 13 compounds (**3** and **4**) showed low-micromolar inhibitory activity in the enzymatic assay. Four more compounds displayed activity in the mid-micromolar range, with similar activities when tested in CHO cells (**5**–**8**). The similar potency of inhibitors **2**–**8** is consistent with the binding mode obtained by docking (Fig. 4), in which the flexible tail is not involved in direct interactions with atoms in the ATP-binding site.

The compounds showing the strongest inhibition effects in the enzymatic assay (**2** and **3** in Table I) were verified in a FRET-based enzymatic assay<sup>41</sup> offered by Cerep (France; see Supplementary Materials). Similar values of  $IC_{50}$  as compared with the original FRET-based enzymatic assay were obtained (Table I).

As a representative of the entire series, inhibitor **2** was investigated further to provide evidence against unspecific binding due to compound aggregation. The Omnia kinetic assay (see Methods section) was performed under two conditions, with and without the addition of detergent (0.01% of Triton X-100). The  $IC_{50}$  values of inhibitor **2** in the two experiments are 5.6 and 7.9  $\mu M$ , which corresponds to a  $K_i$  of 2.7 and 2.9  $\mu M$ , respectively

(Fig. 2). Such detergent-insensitive inhibition is indicative of specific binding.<sup>42</sup>

To further analyze the competitive behaviour of inhibitor **2**, different ATP concentrations were used in the Omnia kinetic assay (ranging from 31.25 to 500  $\mu M$ ). The data obtained in these experiments is shown in a specific velocity plot<sup>38</sup> (Fig. 3), where the ratio of the initial velocities for the noninhibited and inhibited reactions at a given ATP concentration ( $v_0/v_i$  in Fig. 3) is plotted versus  $\sigma/(1+\sigma)$ , with  $\sigma = [ATP]/K_m$ . For an ATP-competitive inhibitor, all curves should intersect close to the point (1,1), which is the case for inhibitor **2**. This result provides strong evidence that inhibitor **2** does indeed bind to the ATP binding site, as predicted by the docking calculations (Fig. 4).

Interestingly, for both compounds **1** and **2**, most of the predicted poses of the anchor fragment and the whole molecule showed significant overlap. In general, small fragments docked in multiple orientations. These docking results are consistent with a recent X-ray study of the binding modes of a 1- $\mu M$  inhibitor of  $\beta$ -lactamase and its constitutive fragments.<sup>43</sup>

### Computational requirements

It is interesting to compare the CPU time required for the preparation and docking of the focused library with docking of all compounds in the three libraries. The ALTA approach required about 6500 h (on a Linux cluster with CPUs with clock speeds of 1.7 GHz): 2 h for decomposition into fragments, 2200 h for fragment docking, 1000 h for substructure search, and 3300 h for flexible-ligand docking and CHARMM minimization.<sup>30</sup> The focused library contains 1/34th of the initial collection of compounds and only about 1/3rd of the fragments, which yields a net speedup by a factor of about 20, a number that would be even larger in the case of flexible-protein docking. Although the actual computation times per compound will naturally vary for other docking programs, the speedup achieved by library preprocessing with the ALTA procedure will remain significant.

## CONCLUSIONS

The number of available small molecules is growing steadily and is already too large for docking with accurate binding energy evaluation and prohibitive for *in vitro* screening. In this article, an automatic fragment-based procedure for focusing libraries of compounds (termed ALTA) was presented and validated. First, molecular fragments are docked and prioritized (i.e., anchors are selected based on a ranking) according to force field energy which includes continuum electrostatics solvation. Large collections of molecules can then be effectively reduced in size by selecting only the compounds for

which one of the constitutive fragments is similar to one of the top ranking anchors.

In principle, ALTA does not require any information about known ligands but in the application presented here pharmacophore knowledge (hydrogen bonds to the hinge region of the receptor tyrosine kinase EphB4) was additionally used to efficiently reduce the size of the original library from about 730,000 to 21,418 molecules. Two series of novel EphB4 inhibitors have been identified in this study. The overall hit rate is very high because only 43 compounds were tested *in vitro*: first 30 molecules suggested by the ALTA approach, and in a second phase 13 compounds with the same anchor (a three-ring system) as the most potent of the seven micromolar inhibitors among the 30 molecules. Compound **2** is a potential candidate for further development for three reasons. First, its low-micromolar affinity and molecular weight of only 353 Da result in a favorable ligand efficiency (binding free energy per heavy atom). Second, the kinetic characterization of compound **2** indicates that it binds to the ATP-binding site, as predicted by the docking calculations. This finding is useful for structure-based lead improvement, for example, by chemical synthesis of derivatives. Finally, derivatives **5–8** have shown inhibitory activity in cell-based assays.

The successful application of the ALTA approach suggests that the identification, based on the structure of the active site, of anchor fragments and their use for library filtering is an efficient strategy for inhibitor discovery by screening *in silico*. Besides its usefulness for high-throughput docking, other possible applications of ALTA include the design of combinatorial libraries and the focusing of large collections of molecules for high-throughput or NMR-based screening *in vitro*.

## ACKNOWLEDGMENTS

The authors thank S. Jelakovic for preparing the model of EphB4, S. Audétat, J. Tietz, P. Schenker, and A. Baici for help with the *in vitro* assays, and C. Bolliger, Th. Steenbock, and A. Godknecht for help with the Matterhorn computer cluster. They also thank A. Widmer (Novartis Pharma, Basel) for providing the molecular modeling program Wit!P which was used for visual analysis. Scripts for ALTA as well as the programs for molecule decomposition into fragments (DAIM), fragment docking (SEED), and flexible-ligand docking (FFLD) are available for free for not-for-profit institutions.

## REFERENCES

- Vangrevelinghe E, Zimmermann K, Schoepfer J, Portmann R, Fabro D, Furet P. Discovery of a potent and selective protein kinase ck2 inhibitor by high-throughput docking. *J Med Chem* 2003;46:2656–2662.
- Jorgensen WL. The many roles of computation in drug discovery. *Science* 2004;303:1813–1818.
- Ding K, Lu Y, Nikolovska-Coleska Z, Qiu S, Ding Y, Gao W, Stuckey J, Krajewski K, Roller PP, Tomita Y, Parrish DA, Deschamps JR, Wang S. Structure-based design of potent non-peptide MDM2 inhibitors. *J Am Chem Soc* 2005;127:10130–10131.
- Cozza G, Bonvini P, Zorzi E, Poletto G, Pagano MA, Sarno S, Donella-Deana A, Zagotto G, Rosolen A, Pinna LA, Meggio F, Moro S. Identification of ellagic acid as potent inhibitor of protein kinase CK2: a successful example of a virtual screening application. *J Med Chem* 2006;49:2363–2366.
- Huang D, Lüthi U, Kolb P, Cecchini M, Barberis A, Caflisch A. In silico discovery of  $\beta$ -secretase inhibitors. *J Am Chem Soc* 2006;128:5436–5443.
- Irwin JJ, Shoichet BK. ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005;45:177–182.
- Kraemer O, Hazemann I, Podjarny AD, Klebe G. Virtual screening for inhibitors of human aldose reductase. *Proteins: Struct Funct Bioinformatics* 2004;55:814–823.
- Lu Y, Nikolovska-Coleska Z, Fang X, Gao W, Shangary S, Qiu S, Qin D, Wang S. Discovery of a nanomolar inhibitor of the human murine double minute 2 (MDM2)-p53 interaction through an integrated, virtual database screening strategy. *J Med Chem* 2006;49:3759–3762.
- Kolb P, Caflisch A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J Med Chem* 2006;49:7384–7392.
- Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins: Struct Funct Genet* 1999;37:88–105.
- Majeux N, Scarsi M, Caflisch A. Efficient electrostatic solvation model for protein-docking. *Proteins: Struct Funct Genet* 2001;42:256–268.
- Budin N, Majeux N, Caflisch A. Fragment-based flexible ligand docking by evolutionary optimization. *Biol Chem* 2001;382:1365–1372.
- Miranker A, Karplus M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins: Struct Funct Genet* 1991;11:29–34.
- Hilpert K, Ackermann J, Banner DW, Gast A, Gubernator K, Hadvary P, Labler L, Müller K, Schmid G, Tschopp T, van de Waterbeemd H. Design and synthesis of potent and highly selective thrombin inhibitors. *J Med Chem* 1994;37:3889–3901.
- Böhm HJ, Böhringer M, Bur D, Gmünder H, Huber W, Klaus W, Kostrewa D, Kühne H, Lübbers T, Meunier-Keller N, Müller F. 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* 2000;43:2664–2674.
- Traxler P, Furet P. Strategies toward the design of novel and selective protein tyrosine kinase inhibitors. *Pharmacol Ther* 1999;82:195–206.
- Furet P, Bold G, Hofmann F, Manley P, Meyer T, Altmann KH. Identification of a new chemical class of potent angiogenesis inhibitors based on conformational considerations and database searching. *Bioorg Med Chem Lett* 2003;13:2967–2971.
- Cheng N, Brantley DM, Chen J. The ephrins and Eph receptors in angiogenesis. *Cytokine Growth Factor Rev* 2002;13:75–85.
- Miyazaki Y, Nakano M, Sato H, Truesdale AT, Stuart JD, Nartey EN, Hightower KE, Kane-Carson L. Design and effective synthesis of novel templates, 3,7-diphenyl-4-amino-thieno and furo-[3,2-c]pyridines as protein kinase inhibitors and in vitro evaluation targeting angiogenic kinases. *Bioorg Med Chem Lett* 2007;17:250–254.
- Milne GWA, Nicklaus MC, Driscoll JS, Wang S, Zaharevitz DW. The NCI drug information system 3D database. *J Chem Inf Comput Sci* 1994;34:1219–1224.
- Liao JL. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J Med Chem* 2007;50:409–424.

22. Cordella L, Foggia P, Sansone C, Vento M. An improved algorithm for matching large graphs. In: Proceedings of the 3rd IAPR TC-15 Workshop on Graph-Based Representations in Pattern Recognition. 2001, Ischia, Italy. pp 149–159.
23. Foggia P. The VFLib Graph Matching Library, version 2.0. <http://amalfi.dis.unina.it/graph/db/vflib-2.0/doc/vflib.html>
24. Wybenga-Groot LE, Baskin B, Ong SH, Tong J, Pawson T, Sicheri F. Structural basis for autoinhibition of the EphB2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region. *Cell* 2001;106:745–757.
25. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680.
26. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815.
27. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Šali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 2000;29:291–325.
28. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
29. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 1993;233:123–138.
30. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* 1983;4:187–217.
31. Momany FA, Rone R. Validation of the general purpose QUANTA 3.2/CHARMm force field. *J Comput Chem* 1992;13:888–900.
32. No K, Grant J, Scheraga H. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. I. Application to neutral molecules as models for polypeptides. *J Phys Chem* 1990;94:4732–4739.
33. No K, Grant J, Jhon M, Scheraga H. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. II. Application to ionic and aromatic molecules as models for polypeptides. *J Phys Chem* 1990;94:4740–4746.
34. Laskowski RA, MacArthur MW, Moss D, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993;26:283–291.
35. Huang D, Caflisch A. Efficient evaluation of binding free energy. *J Med Chem* 2004;47:5791–5797.
36. Kolb P, Huang D, Dey F, Caflisch A. Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J Med Chem* 2008;51:1179–1188.
37. Aparna V, Rambabu G, Panigrahi SK, Sarma JARP, Desiraju GR. Virtual screening of 4-anilinoquinazoline analogues as EGFR kinase inhibitors: importance of hydrogen bonds in the evaluation of poses and scoring functions. *J Chem Inf Model* 2005;45:725–738.
38. Baici A. The specific velocity plot—a graphical method for determining inhibition parameters for both linear and hyperbolic enzyme-inhibitors. *Eur J Biochem* 1981;119:9–14.
39. Sturz A, Bader B, Thierauch KH, Glienke J. EphB4 signaling is capable of mediating ephrinb2-induced inhibition of cell migration. *Biochem Biophys Res Comm* 2004;313:80–88.
40. Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* 2004;9:430–431.
41. Mathis G. Probing molecular interactions with homogeneous techniques based on rare earth cryptates and fluorescence energy transfer. *Clin Chem* 2003;41:1391–1397.
42. Shoichet BK. Screening in a spirit haunted world. *Drug Discov Today* 2006;11:607–615.
43. Babaoglu K, Shoichet BK. Deconstructing fragment-based inhibitor discovery. *Nat Chem Biol* 2006;2:720–723.