

Natural-product-derived fragments for fragment-based ligand discovery

Björn Over^{1,2}, Stefan Wetzel¹, Christian Grütter^{2,3}, Yasushi Nakai^{1†}, Steffen Renner^{1†}, Daniel Rauh^{2,3} and Herbert Waldmann^{1,2*}

Fragment-based ligand and drug discovery predominantly employs sp^2 -rich compounds covering well-explored regions of chemical space. Despite the ease with which such fragments can be coupled, this focus on flat compounds is widely cited as contributing to the attrition rate of the drug discovery process. In contrast, biologically validated natural products are rich in stereogenic centres and populate areas of chemical space not occupied by average synthetic molecules. Here, we have analysed more than 180,000 natural product structures to arrive at 2,000 clusters of natural-product-derived fragments with high structural diversity, which resemble natural scaffolds and are rich in sp^3 -configured centres. The structures of the cluster centres differ from previously explored fragment libraries, but for nearly half of the clusters representative members are commercially available. We validate their usefulness for the discovery of novel ligand and inhibitor types by means of protein X-ray crystallography and the identification of novel stabilizers of inactive conformations of p38 α MAP kinase and of inhibitors of several phosphatases.

Fragment-based ligand and drug discovery (FBDD)¹ allows the exploration of a large fraction of chemical space² with a limited number of low-molecular-weight compounds (molecular weight typically 150–300 g mol⁻¹)³. Most compound libraries used in FBDD are based on privileged substructures delineated from known drugs and drug candidates, and for the most part they therefore cover previously explored chemical space^{4–6}. Despite the positive correlation between both greater complexity and the incorporation of stereogenic centres with the transitioning of compounds from discovery, through clinical testing, and to drugs^{7–9}, these fragments are dominated by sp^2 -rich structures, and three-dimensional fragments remain in high demand¹⁰.

Natural products have played a major role in the discovery of novel ligands and drugs¹¹. They can be regarded as biologically validated, are rich in stereogenic centres and cover segments of chemical space typically not occupied by average synthetic molecules and drugs^{12,13}. An analysis of natural product structures according to or very close to established criteria for fragment likeness^{14,15} could lead to a set of natural-product-derived fragments that distinguish themselves by biological relevance, a high degree of three-dimensionality and access to areas of chemical space as yet largely unexplored by established fragment libraries. Accessibility to and use of such fragments equipped with suitable functional groups for linking would provide a novel opportunity to synthesize natural-product-inspired compound libraries^{16–18} and to overcome limitations in the use of natural products in drug discovery because of a lack of accessibility and synthetic intractability.

Here, we report the cheminformatic analysis of more than 180,000 natural product structures to arrive at 2,000 structurally highly diverse natural-product-derived fragments that are rich in sp^3 -configured centres, distinct from available fragment libraries, and that resemble the properties of natural product scaffolds and of natural products themselves. We demonstrate that the use of such fragments can give access to novel chemotypes for drug

discovery by the identification of structurally new hinge region binders and novel stabilizers of inactive conformations of p38 α MAP kinase, as well as the identification of natural-product-derived fragments for phosphatase inhibitor discovery.

An algorithm for fragment generation

For the fragmentation of natural product structure, the logic and algorithm used previously to establish scaffold trees^{19,20} was modified to retain both attachment points and the chemical nature of the functional groups (to preserve information for later fragment linking and growing) and applied to the Dictionary of Natural Products (DNP)²¹ (Fig. 1 and Supplementary Methods). Thus, directly attached functional groups are kept intact, side chains are shortened, and ring systems are deconstructed in a stepwise manner. Side chains with lengths of up to two atoms are retained, and carbon chains longer than two atoms are pruned after the second atom from the ring. Where the second atom is a heteroatom, pruning will occur after the first following carbon. Carbonyl groups are treated as a single heteroatom. For rings separated by a linker, the connection is pruned and each substructure is then processed separately. The position and type of substitution are therefore retained as in the guiding natural product, and functional groups are differentiated.

To retain, to the greatest extent, the three-dimensional natural product structure, the hybridization and configuration^{19,20} of the centres are not changed during disconnection of fused rings. Each intermediate structure is stored in a primary fragment library, and is then subjected to further filtering. Figure 1 illustrates the process for the natural product Renieramycin P.

Fragment filtering

Fragment generation was performed using the MDL structure data (SD) file of the DNP (DNP 18.2)²¹, which, after standardization, contained 183,769 structures with at least one ring. Fragmentation

¹Max-Planck-Institute of Molecular Physiology, Otto-Hahn-Strasse 11, D-44227 Dortmund, Germany, ²Technische Universität Dortmund, Fakultät Chemie, Chemische Biologie, Otto-Hahn-Strasse 6, D-44227 Dortmund, Germany, ³Chemical Genomics Centre of the Max Planck Society, Otto-Hahn-Strasse 15, D-44227 Dortmund, Germany, [†]Present address: Novartis Institute of Biomedical Research, Basel, Switzerland (S.R.); Shionogi Pharmaceutical Research Center 3-1, Futaba-cho, Toyonaka-shi, Osaka 561-0825, Japan (Y.N.). *e-mail: herbert.waldmann@mpi-dortmund.mpg.de

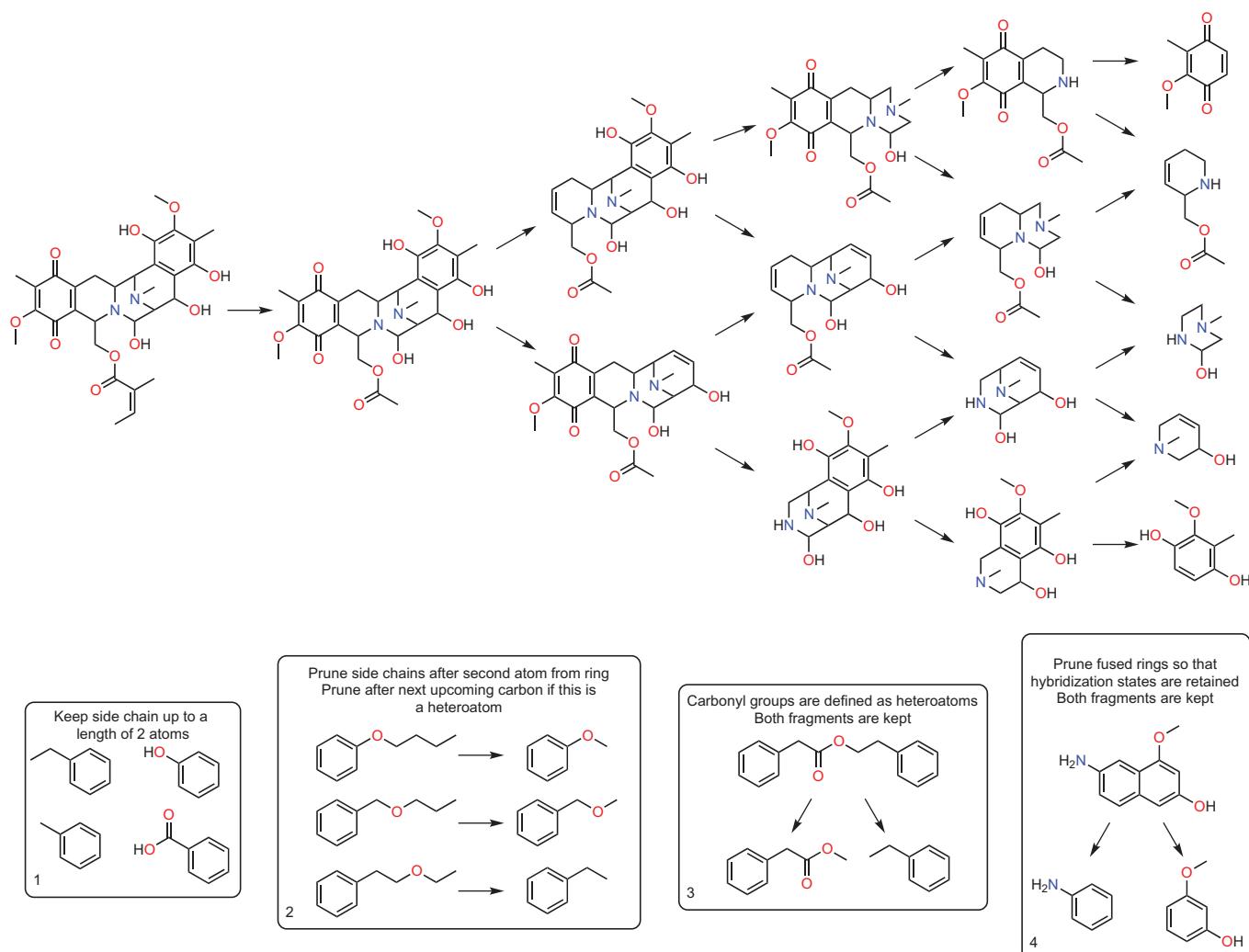


Figure 1 | Fragment generation from the natural product Renieramycin P. The analysis of Renieramycin P provides an example of fragmentation according to the developed algorithm as well as the most important underlying rules. In the first step, side chains are pruned followed by ring degeneration, while keeping all hybridization states. All resulting fragments are passed into the subsequent filtering steps.

yielded 751,577 natural-product-derived fragments (Fig. 2, Supplementary Fig. S1), which were filtered with Pipeline Pilot in three successive steps.

Application of the knowledge-based HTS filter implemented in Pipeline Pilot removed potentially toxic, unstable or undesirable fragments. Subsequent filtering for fragment-like structures applied criteria close to the empirical ‘rule of three’ (molecular weight <300 Da, fewer than three hydrogen bond donors and acceptors, Clog $P < 3$)¹⁴ and also took into account that this rule may not be entirely valid for natural products¹⁵. Thus, as a filter for fragment likeness we chose Alog $P < 3.5$, a molecular weight of 120–350 Da, ≤ 3 hydrogen bond donors, ≤ 6 hydrogen bond acceptors and ≤ 6 rotatable bonds. Filtering with these criteria yielded $\sim 160,000$ fragments.

Further filtering considered that fragments are usually not macrocycles, should not have too many rings or too few heteroatoms, and should not have multi-bridged structures. In this way, the number of natural-product-derived fragments was reduced to 110,485 (Fig. 2). These contained 18,741 different Murcko fragments²², of which $>75\%$ were represented by at most four structures, thereby highlighting the high structural diversity within the natural-product-derived fragment set. Comparison of this set including natural products that fit our fragment criteria with commercially available fragments contained in the ZINC database²³

revealed that only 1,521 of the 73,725 Murcko scaffolds contained in the fragment subset of ZINC overlap with the Murcko scaffolds of the natural-product-derived fragments. Thus, the newly calculated scaffolds introduce numerous new chemotypes beyond those already available commercially.

To further reduce the number of fragments, iterative clustering was performed, thereby adjusting, in a stepwise manner, the number of clusters, cluster size and structure similarity within a cluster, as determined by the average Tanimoto similarity coefficient²⁴ for comparison with the cluster centre. Cluster size was allowed to vary between one and several hundred so as to represent full structural diversity. A final library size of 500–5,000 (ref. 5) was reached by applying a pharmacophor (FCFP_6) fingerprint, which resulted in 2,000 clusters with high similarity within the clusters and a high diversity of the cluster centres (see Supplementary Fig. S2 for statistics).

Properties of natural-product-derived fragments

The distributions of the numbers of stereocentres, hydrogen bond donors and acceptors, as well as other properties of the cluster centres and the natural product fragments obtained after filtering ($\sim 110,000$) compare very well and come close to the distribution determined for all entries in the DNP (Supplementary Fig. S3).

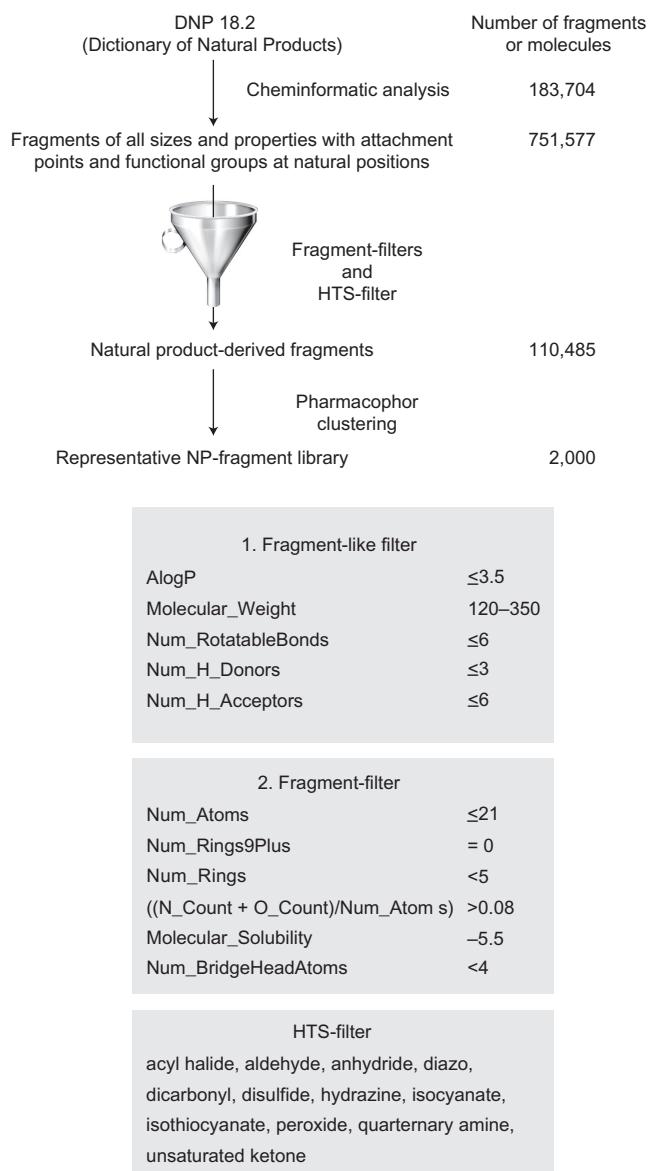
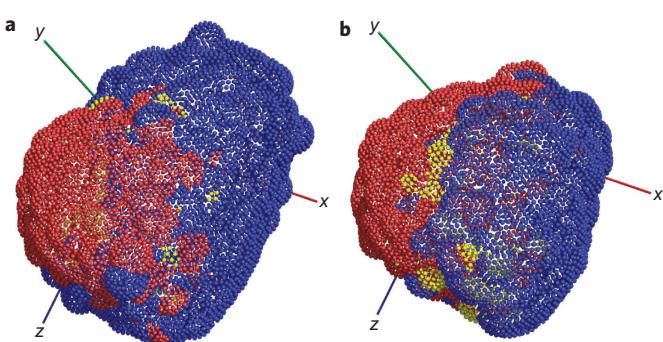


Figure 2 | Filtering and clustering process. Initially, the HTS filter implemented in Pipeline Pilot removes potentially toxic, unstable or undesirable fragments (acyl-halide, aldehyde, anhydride, diazo, dicarbonyl, disulfide, hydrazine, isocyanate, isothiocyanate, peroxide, quaternary amine, and unsaturated ketone). The remaining fragments are then filtered for fragment-like structures according to the following criteria: AlogP < 3.5 , molecular weight of 120–350 Da, ≤ 3 hydrogen bond donors, ≤ 6 hydrogen bond acceptors and ≤ 6 rotatable bonds [fragment-like filter: atom-based prediction of the partition coefficient (AlogP) ≤ 3.5 , molecular_weight of 120–350 Da, number of rotatable bonds and hydrogen bond acceptors (Num_RotatableBonds, Num_H_Acceptor) ≤ 6 , number of hydrogen bond donors (Num_H_Donors) ≤ 3]. Further filtering addresses the occurrence of macrocycles, the number of rings and heteroatoms, and the number of multibridged structures [number of rings with a ring size higher than nine atoms (Num_Rings9Plus) = 0, total number of rings (Num_Rings) < 5 , sum of nitrogen and oxygen atoms divided by the total number of atoms ($((N_Count + O_Count)/Num_Atoms) > 0.08$, molecular_solubility > -5.5 , number of bridgehead atoms (Num_BridgeHeadAtoms) < 4]. Further reduction by iterative clustering according to the average Tanimoto similarity coefficient compared to the cluster centre resulted in 2,000 clusters.

Mapping and extraction of the Murcko scaffolds²² of the cluster centres onto the natural product scaffold tree¹⁹ (Supplementary Fig. S4) revealed that the cluster centre scaffolds cover a large



	NP-fragments	ZINC-fragments	SynLib-fragments
Fsp3_Mean	0.58	0.42	0.41
O_Count_Mean	2.33	1.64	1.46
N_Count_Mean	1.11	2.05	2.62
Hal_Count_Mean	0.08	0.30	0.25
Num_Rings_Mean	2.40	1.67	2.12
Num_AromaticRings_Mean	0.48	1.04	0.97
Num_NonAromaticRings_Mean	1.92	0.63	1.15

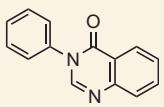
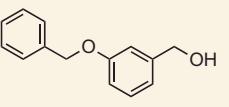
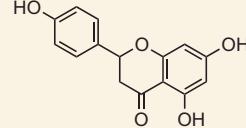
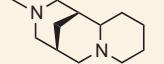
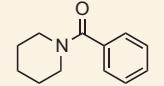
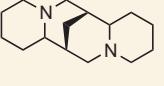
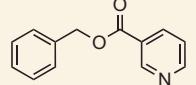
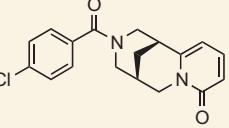
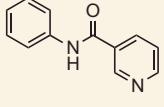
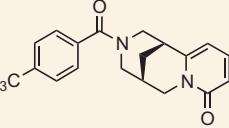
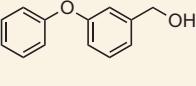
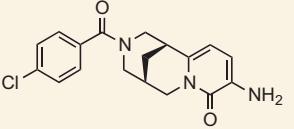
Figure 3 | Comparison of the chemical space defined by the natural product fragments and the chemical space represented by fragments derived from commercially available compounds. Each dot represents a cluster of, on average, 25 fragments with similar property constellations. **a**, PCA of the molecular properties of all generated natural product fragments passing the filters (NpLib) and the ZINC fragment (ZincFrag) subset. Clusters (dots) with two-thirds or more natural product fragment content (NpFrag clusters) are shown in red. Clusters with two-thirds or more zinc fragment content (ZincFrag) are shown in blue. All clusters between are shown in yellow. **b**, PCA of NpLib and SynLib fragments. Clusters with two-thirds or more natural product fragment content (NpFrag clusters) are shown in red. Clusters with two-thirds or more SynLib fragments (SynLib) are shown in blue. All clusters between are shown in yellow. The PCA shows that there is only a small overlap between the chemical spaces defined by the natural-product-derived fragment library generated in this work and either of the synthetic fragment libraries (ZincFrag or SynLib). Statistics for mean counts of relevant parameters are given below the graphic depiction: Fsp3 = number of sp^3 -hybridized carbons/total carbon count⁹, O/N/Hal_Count = number of oxygen/nitrogen/halogen atoms respectively, Num_Rings = total number of rings, Num_AromaticRings = number of aromatic rings, Num_NonAromaticRings = number of non-aromatic rings.

fraction of the natural product tree. The non-covered structures are predominantly α,β -unsaturated ketones, which were filtered out by the HTS filter (see above) because of their reactivity.

Comparison of the chemical space defined by the natural product fragments with the chemical space represented by the six times larger commercially available fragment subset annotated in the ZINC database (zinc.docking.org) by means of principal component analysis (PCA; Fig. 3a) showed that these two fragment sets overlap only to a minor extent (see above) and have substantially different properties (Supplementary Methods and Fig. S5). In a second analysis, a collection of 247,000 compounds was compiled by random choice from various vendor libraries. This was fragmented and analysed in analogy to the natural products to yield a fragment library with a set of structures representative of synthetic molecules (SynLib; for the distribution of properties see Supplementary Fig. S6).

After PCA, all generated fragments were grouped into clusters of compounds with the same property constellation. Three-dimensional graphical representation using the first principal components as parameters revealed that the three clusters are clearly different and represent distinct areas of chemical space (Fig. 3; for property distribution see Supplementary Figs S3, S5 and S6).

Table 1 | Inhibition of p38 α MAP kinase by the identified natural product fragments and derivatives thereof.

Entry	Fragment	IC ₅₀ (μ M)*	LE	Entry	Fragment	IC ₅₀ (μ M)	LE
1		692 ± 252	0.24	7		6,247 ± 358	0.18
2		292 ± 54	0.23	8		1,341 ± 215	0.28
3		1,534 ± 128	0.27	9		2,234 ± 445	0.21
4		818 ± 80	0.26	10		5,103 ± 544	0.13
5		3,471 ± 796	0.22	11		5,453 ± 513	0.13
6		520 ± 192	0.29	12		435 ± 138	0.19

*All IC₅₀ values were calculated from at least three independent measurements and are given with standard deviation.

Natural product-derived p38 α inhibitors

For the experimental demonstration that natural-product-derived fragments can yield hitherto unexplored ligand classes and areas of chemical space for established drug targets, p38 α MAP kinase was chosen²⁵. A search of the ZINC database (www.zinc.docking.org), ChemNavigator (www.chemnavigator.com) and eMolecules (www.emolecules.com) revealed that ~5% of the cluster centres are commercially available (106). However, a consideration of all cluster members revealed that representative natural-product-derived fragments are commercially available for nearly half (872) of the 2,000 clusters. The most representative members of the 872 clusters, including all 106 available cluster centres (for 17 clusters only the cluster centre is available; Supplementary Table S1), displayed the highest Tanimoto index (FCFP_2 or FCFP_6) compared to the cluster centre, with a minimum Tanimoto of >0.5 (on average 0.75 for FCFP_2 and 0.61 for FCFP_6; Supplementary Fig. S7).

In total, we obtained or synthesized 193 cluster centres, near cluster members or closely related commercially available stable analogues (for representative examples see Tables 1 and 2; the whole set of selected fragments is shown in Supplementary Fig. S8). An example for a closely related fragment is **8**, which is structurally related to the centre of cluster 509, with 85 members, and is a substructure of 53 natural products, such as the alkaloid sparteine (Table 1, Supplementary Fig. S9). The choice of such close analogues was indicated in particular when the

chemoinformatic analysis yielded unstable fragments like enamines. Often, slight modifications such as acylation were sufficient for the generation of a stable closely related analogue, and in various cases inspiration could be drawn from inspection of the guiding natural product.

For nine fragments, crystal structures complexed with p38 α MAP kinase were determined (Fig. 4 and Supplementary Figs S10 and S11).

Quinazolinone fragment **1** binds to Met109 in the hinge region of the kinase via a water-mediated hydrogen bond (Fig. 4a), while the attached aromatic ring locks the fragment in a hydrophobic sub-pocket in the back of the ATP binding site. This binding mode was only very recently identified independently for related compounds²⁶. The binding and orientation of flavanone **2** in the binding site are similar (Fig. 4b). Flavones and flavanones have been identified previously as kinase inhibitors²⁷. Hinge region-binding fragments **3**, **4** and **5** (Fig. 4c–e) represent new fragments for kinase inhibitor development. Recently, related fragments with similar potency (IC₅₀ = 1.3 mM and 778 μ M, respectively) were identified in an independent FBDD effort^{25,26}.

Fragments **6** and **7** bind p38 α MAP kinase differently. Whereas **6** bridges the hydrophobic pocket and the hinge region, and its benzyl alcohol substructure forms a hydrogen bond to the backbone of Met109 (Fig. 4f), fragment **7** does not contact the hinge region and binds in a reversed manner with the benzyl alcohol pointing towards the DFG motif (Fig. 4g). In this case, a

Table 2 | Inhibition of representative phosphatases by the identified natural product fragments and derivatives thereof.

Entry	Fragment	Phosphatase*	IC_{50} (μM) [†]	LE	Entry	Fragment	Phosphatase	IC_{50} (μM)	LE
13		VEPTP	16.2 ± 3.5	0.49	18		SHP2	71.1 ± 10.6	0.51
14		VHR	68.2 ± 23.2	0.56	19		PTP1B	20.9 ± 7.7	0.48
15		CDC25A	10.7 ± 0.8	0.66	20		CDC25A	86.0 ± 7.1	0.50
16		MPTPB	80.8 ± 9.4	0.45	21		VEPTP	91.50 ± 5.21	0.49
17		VEPTP	25.5 ± 5.2	0.44					

*As representative phosphatases, *Mycobacterium tuberculosis* protein tyrosine phosphatases A and B (MptpA, MptpB), vascular endothelial protein tyrosine phosphatase (V-PTP), cell division cycle 25 homologue A (CDC25A), VH1-related phosphatase (VHR), SH2 domain containing phosphatase (SHP2) and the protein tyrosine phosphatase 1B (Ptp1B) were investigated.

[†]All IC_{50} values were calculated from at least three independent measurements and are given with standard deviation.

second fragment is located in the allosteric pocket, reminiscent of a type III inhibitor.

Additional activity assays identified the structurally truly novel sp^3 -configured fragments **8** and **9** as weak inhibitors of p38 α MAP kinase (Table 1), for which the co-crystal structure indicated binding to the inactive DFG-out conformation of p38 α MAP kinase. Subsequent synthesis of a collection of cytisine and sparteine derivatives following published procedures²⁸ and based on the structures of the centres of clusters 29, 509 and 1,846 (Supplementary Figs S12, S13 and S14) and subsequent crystal structure determination revealed that cytisine derivatives **10** and **11** indeed bind to the allosteric pocket of p38 α as a novel class of type III inhibitors. These ligands stabilize the DFG motif of the kinase in the enzymatically inactive DFG-out or DFG-in-between conformation (Fig. 4*h,i*). These findings are of particular interest because allosteric inhibitors²⁹, as a result of their unique binding characteristics (for example, slow off-rate), are one focus of modern kinase inhibitor research³⁰. Optimized type III inhibitors, which bind exclusively to the allosteric site, allow the inhibitor to grow outside the active site, and therefore explore less conserved sites to increase kinase inhibitor selectivity^{31,32}. Only a very limited number of type III kinase inhibitors are known, most clustering around the pyrazole urea scaffold derived from the well-studied type II inhibitor BIRB-796 (Fig. 4*j*). Thus, the unique molecular shape of sp^3 -rich fragments **10** and **11** offers unprecedented starting points for further kinase inhibitor research. The crystal structure of **10** in complex with the kinase also revealed that two additional molecules of compound **10** bound to the protein surface with 50% occupancy, sandwiching Trp196 (Supplementary Fig. S11). In the crystal structure including the closely related cytisine derivative **11**, such additional binding was not observed, indicating that the allosteric binding mode represents the main interaction.

Inhibitory activity was determined by means of an activity-based kinase assay (Supplementary Methods), and fragments were included that had shown inconclusive electron density in the crystallization experiments. The inhibition data (Table 1) obtained for the hinge region binders are in the medium range for fragment-type compounds. The allosteric inhibitors display relatively low IC_{50} values and ligand efficiencies, which may be due to the fact that they were identified in an activity-based assay originally developed for the identification of inhibitors targeting the ATP binding site. In addition to the fragments described above, sparteine **9** and sparteine analogue **8** were identified as novel sp^3 -configured natural-product-derived kinase inhibitors. Notably, the introduction of an amino group at the 9-position (compound **12**) increased the potency of the cytisine derivative **10** by one order of magnitude. Although no complex crystal structure could be obtained for **12**, modelling studies suggest that the observed significant increase in activity could result from additional polar interactions of the primary amine with Arg67 of α -helix C and the side chain of Glu71, two amino acids that are crucial for binding type III ligands and control access to the allosteric or switch pocket^{31,33}.

Natural product-derived phosphatase inhibitors

To demonstrate that natural-product-derived fragments may also yield hits for relevant target classes that have proven difficult to address, we screened for inhibitors of several representative tyrosine- and dual-specificity phosphatases and included in the screen natural product fragment cluster members available in house from our previous efforts in biology oriented synthesis (BIOS)^{16–18,34}. These enzymes have been the subject of a series of medicinal chemistry programmes, but inhibitor development has proven difficult^{35–38}.

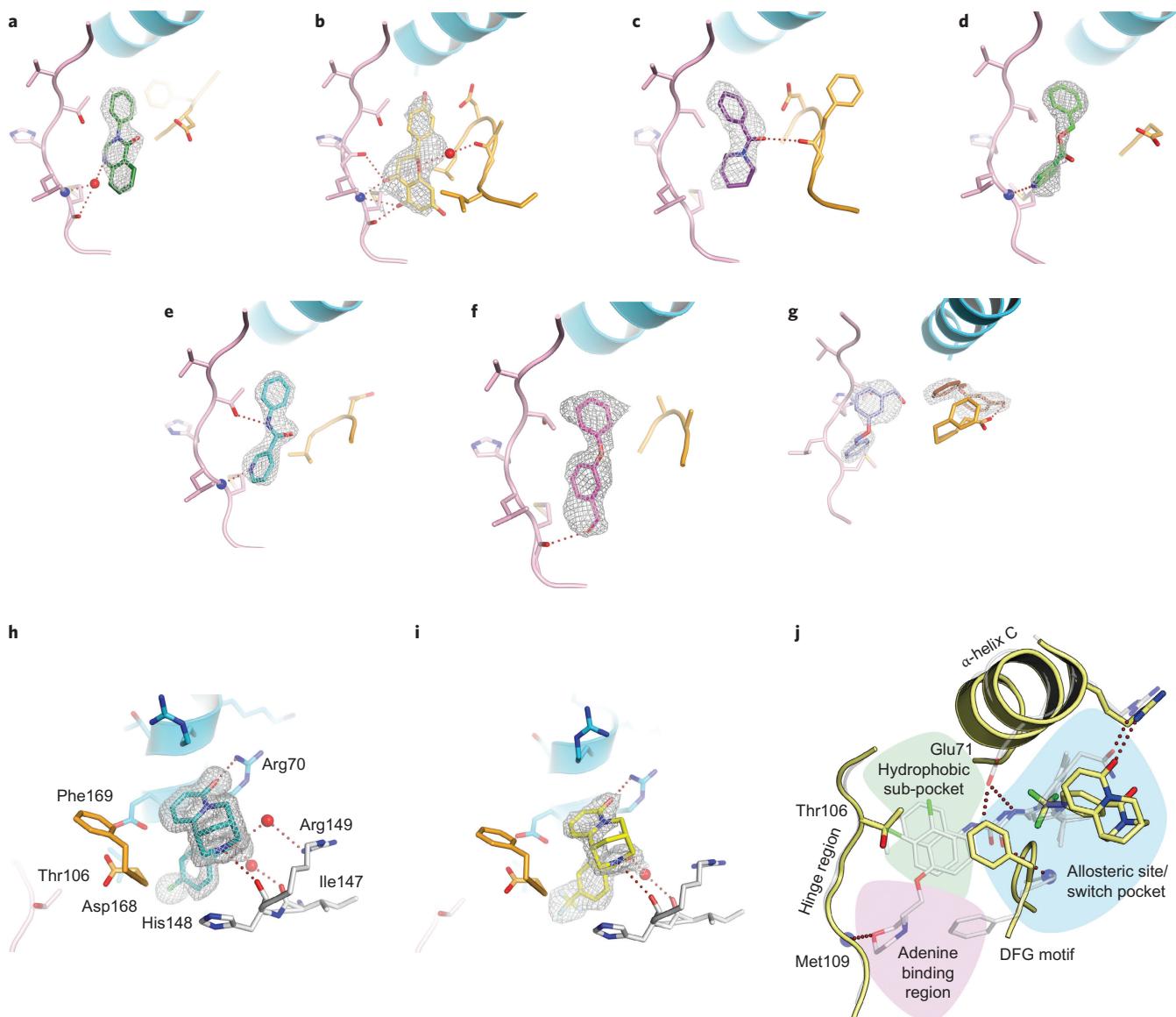


Figure 4 | Crystal structures of p38 α MAP kinase in complex with natural-product-derived fragments. Electron density maps ($2\text{Fo}-\text{Fc}$) are contoured at 1σ . The hinge region (pink), α -helix-C (blue), DFG-motif (orange) and polar interactions (red dotted lines) are highlighted. **a, b**, Quinazolinone **1** (**a**) forms a water-mediated hydrogen bond to the hinge backbone (Met109), whereas the natural product naringenin **2** has a direct interaction (**b**). **c-f**, Fragments **3** (**c**), **4** (**d**), **5** (**e**) and **6** (**f**) show comparable binding modes addressing the hydrophobic sub-pocket next to the gatekeeper (Thr106). **g**, Fragment **7** is partially defined at the hinge region and occupies the allosteric pocket. **h, i**, The cytisine derivatives **10** (**h**) and **11** (**i**) exclusively occupy the allosteric pocket of p38 α . **j**, Alignment of the p38 α -**11** complex (yellow) with BIRB-796 (1KV2) and a type III inhibitor (3NNW, shown in white). Both type II and III ligands form hydrogen bonds to Glu71 (α -helix C), the DFG motif, and extend partially into the hinge-binding region. In contrast, the newly identified sp^3 -rich fragment **11** adopts a unique, yet unrepresented binding mode by solely residing in the allosteric pocket. All structures have been deposited (under PDB-ID codes 4EH2, 4EH3, 4EH4, 4EH5, 4EH6, 4EH7, 4EH8, 4EH9 and 4EHV). For additional information see Supplementary Figs S10 and S11.

Inhibition data for cluster members **13–15** and close derivatives of cluster centres (**16–21**) are given in Table 2. A full list of the 52 fragments and small molecules identified with $\text{IC}_{50} < 100 \mu\text{M}$ is given in Supplementary Tables S2 and S3. The results demonstrate that the natural-product-derived fragment set successfully yielded inhibitors with chemotypes structurally distinct from established phosphatase inhibitor classes for all phosphatases investigated. Fragments **13–21** display appreciable IC_{50} values and ligand efficiencies (LE; average LE for all shown phosphatase inhibitors = 0.36).

Discussion

Biological relevance is a key criterion to be met by compounds used in the discovery of protein ligands and inhibitors^{16–18,39}. This

criterion is met in particular by natural products selected by evolution and does, by analogy, also apply to the fragments collectively representing the chemical space defined by the structures of natural products.

To identify natural-product-derived fragments we developed a chemoinformatic method that sequentially deconstructs natural product structures by analogy to the procedure employed for hierarchical arrangement and analysis of compound scaffolds, based on the logic used for the establishment of scaffold trees reported earlier^{19,20}. Application of the algorithm and subsequent filtering using criteria that are a close analogy of the ‘rule of three’, as well as clustering of the obtained fragments according to Tanimoto similarity of pharmacophore fingerprints, led to a library of 2,000 natural-product-derived fragments. This library is

structurally highly diverse, and resembles the distribution of natural product fragment properties and the natural products themselves (for example, oxygen and nitrogen atom content, number of hydrogen bond donors and acceptors, and number of non-aromatic rings). It is rich in sp^3 -configured centres. In the process of fragment generation, attachment points and type of substituent of the fragments are conserved, representing sites for attachment of additional fragments in the course of fragment linking and growth. This library covers an area of chemical space that differs markedly from the chemical space defined by commercially available fragments and fragments generated by analogy from non-natural commercially available compounds.

Because of these characteristics, the use of natural-product-derived fragments in fragment-based ligand discovery promises to give access to structurally novel inhibitor and ligand types for established drug targets. In particular, natural-product-derived fragments may show novel paths to sp^3 -rich compounds that are not readily accessible from established fragment libraries rich with sp^2 -configured molecules.

We provide proof-of-principle for this notion by identifying bicyclic cytisine/sparteine derivatives as unprecedented fragment-sized inhibitors of p38 α MAP kinase binding to the allosteric pocket of the enzyme, and by the identification of several natural-product-derived fragments that may inspire the development of novel phosphatase inhibitor classes.

For nearly half of the clusters, representative members are available commercially. Hert *et al.* found that 83% of the core ring scaffolds present in natural products are absent from the available molecules and screening libraries, and concluded that an enrichment of screening libraries with biogenic compounds is highly desirable⁴⁰. Meanwhile, we have determined that 93% of the Murcko scaffolds of the natural-product-derived fragments are not contained within the Murcko scaffolds of the fragment subset of the ZINC database. Thus, the investigations are mutually supportive.

The natural-product-derived fragments are often far less complex structurally than the guiding natural products themselves (Fig. 1). However, their synthesis will often still require considerable synthetic effort, and for widespread access to the full set of natural-product-derived fragments, the development of novel, efficient synthesis methodologies is required. However, the syntheses of natural-product-derived fragments will by no means have to meet the level of difficulty encountered in the multi-step synthesis of genuine natural products. In particular, by retaining their linker attachment points, many of the fragments have chemical handles that allow for efficient growth of fragment hits into larger molecules using established and robust chemical reactions, while retaining the structural complexity of the natural products. Thus, the use of natural-product-derived fragments in fragment-based design defines a novel approach to the syntheses of natural-product-inspired compound collections and may serve to overcome the limitations in the use of natural products in drug discovery due to lack of accessibility and synthetic tractability.

Fragments resulting from the clustering and filtering procedure may define chemically unstable molecules such that alteration of the structure may be required to arrive at compounds suitable for screening and inhibition studies. Such changes may include equipment of a cluster centre with additional substituents (for example, acylation of an enamine) or consideration of a structurally closely related yet chemically stable cluster member. Such a hopping within clusters⁴¹ or fragment shuffling within evolved inhibitors⁴² appears justified, because fragments with a high overlap of their three-dimensional structure and thereby pharmacophore similarity often display similar inhibition profiles⁴³.

The method used here for the identification of natural product fragments is conceptually based on and parallels the structural

classification of natural products that led to the hierarchical natural product tree (see above). Very recently, the scaffold tree approach has been extended to scaffold network generation⁴⁴. We note that this extended branch ramification considering approach, by analogy, should be applicable to the fragments generated in the deconstruction of natural products.

Methods

Fragment generation algorithm. The fragment generation algorithm analyses and simplifies structures by analogy to the scaffold tree generation developed previously^{19,20,39}. The actual algorithm was implemented as a modified version of the **ScaffoldTreeGenerator**, which is part of the **Scaffold Hunter software**⁴⁵. The cheminformatics part was implemented using the Chemistry Development Kit (CDK)⁴⁶. The fragmentation algorithm starts with extraction of the Murcko scaffold decorated with attachment points from each molecule. Subsequently, rings and connecting linker chains are removed from each decorated scaffold. Chains are only removed to the next attachment point. This process is continued until all rings are removed. All generated fragments are stored in a database.

Pseudocode:

```
For each Molecule
    Identify Murcko Scaffold
    Keep attachment points
    Re-do while rings can be removed
    Generate parent fragments by removing each terminal ring once and cut
    back the linker to the attachment point
    Use parent fragments as new starting points
```

Attachment points are chemically modifiable groups defined by hetero atoms in the side chain. The program was designed to retain functional groups in particular, and only cut at carbon–carbon bonds. Hence, the initial functional motif remains intact in the fragment and can be identified. The program then generates all possible fragments by removal of one ring (and its link) at a time.

The algorithm detects attachment points by traversing through the side chain through two atoms, starting from the ring-based aliphatic atom. For each atom it decides—based on a set of rules (Supplementary Scheme S1)—whether the atom is kept, and whether the neighbour atoms are examined. The rules decide whether this examination occurs only if the neighbour atoms are within the distance limit from the ring, or in every case. The algorithm therefore keeps aliphatic carbons as long as they are within two atoms from the ring. If the algorithm encounters a carbon atom linked to a non-carbon and non-hydrogen atom, it will keep it and examine only the immediate neighbour atom, irrespective of its distance from the ring.

If a branching atom is a carbon atom, the program continues to transverse the branches separately, within the distance limit. If the branching carbon is connected to a heteroatom by a double bond, the immediate next neighbour atoms will be investigated, irrespective of their distance from the ring. These rules allow functional groups such as carbonyl groups or esters to remain intact. All other carbons are kept (as terminating carbon) and the rest of the chain is removed. If the branching atom happens to be a non-carbon and non-hydrogen atom, the immediate next neighbour atoms will be investigated, irrespective of their distance from the ring. This part of the algorithm is visualized in the flow chart in Supplementary Scheme S1.

Received 24 April 2012; accepted 19 October 2012;
published online 2 December 2012

References

- Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nature Chem.* **1**, 187–192 (2009).
- Roughley, S. D. & Hubbard, R. E. How well can fragments explore accessed chemical space? A case study from heat shock protein 90. *J. Med. Chem.* **54**, 3989–4005 (2011).
- Chessari, G. & Woodhead, A. J. From fragment to clinical candidate—a historical perspective. *Drug Discov. Today* **14**, 668–675 (2009).
- Babaoglu, K. & Shoichet, B. K. Deconstructing fragment-based inhibitor discovery. *Nature Chem. Biol.* **2**, 720–723 (2006).
- Hubbard, R. E., Chen, I. & Davis, B. Informatics and modeling challenges in fragment-based drug discovery. *Curr. Opin. Drug Discov. Dev.* **10**, 289–297 (2007).
- Siegel, M. G. & Vieth, M. Drugs in other drugs: a new look at drugs as fragments. *Drug Discov. Today* **12**, 71–79 (2007).
- Nicholls, A. *et al.* Molecular shape and medicinal chemistry: a perspective. *J. Med. Chem.* **53**, 3862–3886 (2010).
- Ritchie, T. J., Macdonald, S. J. F., Young, R. J. & Pickett, S. D. The impact of aromatic ring count on compound developability: further insights by examining carbo- and hetero-aromatic and -aliphatic ring types. *Drug Discov. Today* **16**, 164–171 (2011).

9. Lovering, F., Bikker, J. & Humblet, C. Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **52**, 6752–6756 (2009).
10. Hung, A. W. *et al.* Route to three-dimensional fragments using diversity-oriented synthesis. *Proc. Natl Acad. Sci. USA* **108**, 6799–6804 (2011).
11. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **75**, 311–335 (2012).
12. Wetzel, S., Schuffenhauer, A., Roggo, S., Ertl, P. & Waldmann, H. Chemoinformatic analysis of natural products and their chemical space. *Chimia Int. J. Chem.* **61**, 355–360 (2007).
13. Grabowski, K., Baringhaus, K.-H. & Schneider, G. Scaffold diversity of natural products: inspiration for combinatorial library design. *Nat. Prod. Rep.* **25**, 892–904 (2008).
14. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A ‘Rule of Three’ for fragment-based lead discovery? *Drug Discov. Today* **8**, 876–877 (2003).
15. Köster, H. *et al.* A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.* **54**, 7784–7796 (2011).
16. Wetzel, S., Bon, R. S., Kumar, K. & Waldmann, H. Biology-oriented synthesis. *Angew. Chem. Int. Ed.* **50**, 10800–10826 (2011).
17. Bon, R. S. & Waldmann, H. Bioactivity-guided navigation of chemical space. *Acc. Chem. Res.* **43**, 1103–1114 (2010).
18. Kumar, K. & Waldmann, H. Synthesis of natural product inspired compound collections. *Angew. Chem. Int. Ed.* **48**, 3224–3242 (2009).
19. Schuffenhauer, A. *et al.* The scaffold tree: visualization of the scaffold universe by hierarchical scaffold classification. *J. Chem. Inf. Model.* **47**, 47–58 (2007).
20. Koch, M. A. *et al.* Charting biologically relevant chemical space: a structural classification of natural products (SCONP). *Proc. Natl Acad. Sci. USA* **102**, 17272–17277 (2005).
21. *Dictionary of Natural Products* Version 18.2, 2009 (Chapman & Hall/CRC Press, 2009).
22. Bemis, G. W. & Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **39**, 2887–2893 (1996).
23. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **52**, 1757–1768 (2012).
24. Butina, D. Unsupervised data base clustering based on daylight’s fingerprint and Tanimoto similarity: a fast and automated way to cluster small and large data sets. *J. Chem. Inf. Comput. Sci.* **39**, 747–750 (1999).
25. Gill, A. L. *et al.* Identification of novel p38 α MAP kinase inhibitors using fragment-based lead generation. *J. Med. Chem.* **48**, 414–426 (2004).
26. Pollack, S. *et al.* A comparative study of fragment screening methods on the p38 α kinase: new methods, new insights. *J. Comput. Aided Molec. Des.* **25**, 677–687 (2011).
27. Goettert, M., Schattel, V., Koch, P., Merfort, I. & Laufer, S. Biological evaluation and structural determinants of p38 α mitogen-activated-protein kinase and c-Jun-N-terminal kinase 3 inhibition by flavonoids. *ChemBioChem* **11**, 2579–2588 (2010).
28. Dearden, M. J., McGrath, M. J. & O’Brien, P. Evaluation of (+)-sparteine-like diamines for asymmetric synthesis. *J. Org. Chem.* **69**, 5789–5792 (2004).
29. Simard, J. R. *et al.* A new screening assay for allosteric inhibitors of cSrc. *Nature Chem. Biol.* **5**, 394–396 (2009).
30. Rabiller, M. *et al.* Proteus in the world of proteins: conformational changes in protein kinases. *Arch. Pharm. (Weinheim)* **343**, 193–206 (2010).
31. Ahn, Y. M. *et al.* Switch control pocket inhibitors of p38-MAP kinase. Durable type II inhibitors that do not require binding into the canonical ATP hinge region. *Bioorg. Med. Chem. Lett.* **20**, 5793–5798 (2010).
32. Swann, S. L. *et al.* Biochemical and biophysical characterization of unique switch pocket inhibitors of p38 α . *Bioorg. Med. Chem. Lett.* **20**, 5787–5792 (2010).
33. Getlik, M. *et al.* Structure-based design, synthesis and biological evaluation of N-pyrazole, N'-thiazole urea inhibitors of MAP kinase p38 α . *Eur. J. Med. Chem.* **48**, 1–15 (2012).
34. Kaiser, M., Wetzel, S., Kumar, K. & Waldmann, H. Biology-inspired synthesis of compound libraries. *Cell. Mol. Life Sci.* **65**, 1186–1201 (2008).
35. Vintonyak, V. V., Antonchick, A. P., Rauh, D. & Waldmann, H. The therapeutic potential of phosphatase inhibitors. *Curr. Opin. Chem. Biol.* **13**, 272–283 (2009).
36. Vintonyak, V. V., Waldmann, H. & Rauh, D. Using small molecules to target protein phosphatases. *Bioorg. Med. Chem.* **19**, 2145–2155 (2011).
37. Bialy, L. & Waldmann, H. Inhibitors of protein tyrosine phosphatases: next-generation drugs? *Angew. Chem. Int. Ed.* **44**, 3814–3839 (2005).
38. Nören-Müller, A. *et al.* Discovery of protein phosphatase inhibitor classes by biology-oriented synthesis. *Proc. Natl Acad. Sci. USA* **103**, 10606–10611 (2006).
39. Renner, S. *et al.* Bioactivity-guided mapping and navigation of chemical space. *Nature Chem. Biol.* **5**, 585–592 (2009).
40. Hert, J., Irwin, J. J., Laggner, C., Keiser, M. J. & Shoichet, B. K. Quantifying biogenetic bias in screening libraries. *Nature Chem. Biol.* **5**, 479–483 (2009).
41. Hessler, G. & Baringhaus, K.-H. The scaffold hopping potential of pharmacophores. *Drug Discov. Today: Technologies* **7**, e263–e269 (2010).
42. Nisius, B. & Rester, U. Fragment shuffling: an automated workflow for three-dimensional fragment-based ligand design. *J. Chem. Inf. Model.* **49**, 1211–1222 (2009).
43. Bamborough, P., Brown, M. J., Christopher, J. A., Chung, C.-W. & Mellor, G. W. Selectivity of kinase inhibitor fragments. *J. Med. Chem.* **54**, 5131–5143 (2011).
44. Varin, T., Schuffenhauer, A., Ertl, P. & Renner, S. Mining for bioactive scaffolds with scaffold networks: improved compound set enrichment from primary screening data. *J. Chem. Inf. Model.* **51**, 1528–1538 (2011).
45. Wetzel, S. *et al.* Interactive exploration of chemical space with scaffold hunter. *Nature Chem. Biol.* **5**, 581–583 (2009).
46. Steinbeck, C. *et al.* Recent developments of the chemistry development kit (CDK)—an open-source Java library for chemo- and bioinformatics. *Curr. Pharmaceut. Des.* **12**, 2111–2120 (2006).

Acknowledgements

The research leading to these results was supported by funding from the European Research Council (ERC) under the European Union’s Seventh Framework Program (FP7/2007–2013) and ERC grant agreement no. 268309, from the German Federal Ministry for Education and Research through the German National Genome Research Network-Plus (NGFN-Plus) (grant no. BMBF 01GS08104 to H.W. and D.R.), as well as from the Fonds der Chemischen Industrie.

Author contributions

B.O. and Y.N. performed computational experiments and syntheses. S.W. and S.R. performed computational experiments. B.O. performed biochemical experiments. B.O. and C.G. determined crystal structure analyses. B.O., S.W., D.R. and H.W. designed experiments. D.R. and H.W. supervised the research. B.O., S.W., D.R. and H.W. wrote the manuscript.

Additional information

Supplementary information and chemical compound information are available in the online version of the paper. Reprints and permission information is available online at <http://www.nature.com/reprints>. Correspondence and requests for materials should be addressed to H.W.

Competing financial interests

The authors declare no competing financial interests.