FRAGMENT-BASED LEAD DISCOVERY

David C. Rees, Miles Congreve, Christopher W. Murray and Robin Carr

Fragment-based lead discovery is gaining momentum in both large pharmaceutical companies and biotechnology laboratories as a complementary approach to traditional screening. This is because fragment-based approaches require significantly fewer compounds to be screened and synthesized, and are showing a high success rate in generating chemical series with lead-like properties. Compared with traditional screening hits, the starting fragments have considerably lower molecular mass, and although the binding interactions of these fragments with a target protein are weak, they are structurally understood through X-ray crystallography or NMR, and they exhibit high 'ligand efficiency'. Here, we use examples from 25 different protein targets to describe chemical strategies that exploit this structural knowledge to rapidly develop fragments into high-affinity leads.

The concept of a 'target-rich, lead-poor' pipeline in drug discovery, and widespread concern about the attrition rate of chemical compounds in (pre)clinical development, are together fuelling the search for better quality hits and chemical lead series. Researchers are rising to this challenge by devising new ways to identify chemical leads for specific protein targets and by using as starting points chemical structures that reflect the physical properties of successful oral drug molecules. A particular approach to lead identification for drug discovery involves the selection, screening and optimization of so-called fragments (also referred to as needles[1], shapes[2], binding elements[3] or seed templates[4]). This review discusses fragment-based lead discovery, and focuses on the output of this new approach by collating published examples from 25 protein targets. These targets are primarily enzymes, and the screening techniques used include X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, *in vitro* bioassays and mass spectrometry.

The concepts that underpin the chemical fragments approach can be traced back to the pioneering work of Jencks[5] and Ariens[6], who showed that drug-like molecules can be regarded as the combination of two or more individual binding epitopes (or fragments). Screening for these binding epitopes using low-molecular-mass molecules is the essence of the fragments approach. The screened fragments are smaller (typically with M_r = 120–250 or 8–18 non-hydrogen atoms), have less functionality and are correspondingly weaker than most hits from high-throughput screening (HTS), with typical binding affinities in the range mM–30 μM . Accordingly, biophysical screening methods, including NMR[7,8] and X-ray crystallography[9,10], are particularly suitable, given the range of binding affinities that they can detect. In addition, these methods can afford significant structural understanding of the ligand–protein binding event, which is crucial in prioritizing fragment hits and rapidly developing them into leads. This process is illustrated in FIG. 1.

HTS is currently the established approach for hit identification within the pharmaceutical industry. Despite many successes from HTS, there is a need to find alternative approaches to lead discovery. Fragment-based approaches offer a number of attractive features compared with HTS. First, the number of compounds typically screened is in the range of only a hundred to a few thousand, because lower complexity compounds (fragments) have a higher probability of matching a target protein-binding site[11]. Second, a high proportion of the atoms in a fragment hit are directly involved in the desired protein-binding interaction, so fragments can be described as efficient binders (that is, high binding energies per unit molecular mass); generally in HTS, larger and more potent compounds are identified, but the compounds are less efficient binders. Third, when the binding interaction is structurally validated and understood, the subsequent chemical optimization

Astex Technology, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, UK.
Correspondence to D.C.R.
e-mail: d.rees@astex-technology.com
doi:10.1038/nrd1467

benefits from extensive design, synthesis of only a few compounds and a high success rate. And last, starting the chemical optimization stage with a low-molecular-mass fragment is likely to produce leads whose M_r is still within the range desired for lead-likeness[12–14]. FIGURE 2 schematically illustrates how fragments and HTS hits compare as starting points for drug discovery (see also FIG. 3), and TABLE 1 gives more detail on the differences between fragment-based lead discovery and HTS.

Drug-like, lead-like and fragment-like

At present, there are several guidelines for defining drug-like properties[14,15], such as the 'Lipinski RULE OF FIVE'[16], which is used to maximize an oral drug candidate's probability of surviving development, and a more recent analysis based on the number of rotatable bonds, which indicated an upper limit of seven rotatable bonds in orally bioavailable drugs[17].

Although these guidelines are useful for assessing the risk profile of an oral *drug* candidate entering development, they are not necessarily as relevant for assessing the optimum properties of a *lead*. For example, studies of 450 pairs of commercial drugs and their corresponding leads indicated that, on average, leads had lower M_r, lower lipophilicity (cLogP), fewer aromatic rings and fewer hydrogen-bond acceptors[11]. A similar analysis concluded that libraries of compounds with M_r = 100–350 and cLogP = 1–3 are superior for finding leads compared with those comprising drug-like compounds, with higher M_r and cLogP. The reason for this is that current lead-optimization practices routinely increase both M_r (on average by ~80) and lipophilicity (on average more than 1 Log unit) over those of initial leads. So, if the initial lead already possesses drug-like physical properties, then the optimization process is likely to result in drug candidates with poorer drug-like properties. Overall, this suggests that 'small is beautiful' in quality hits and leads[12,13]. BOX 1 describes the results from an analysis of fragment hits against a range of targets, which imply that a 'rule of three' might be useful when constructing fragment libraries[18].

The remainder of this review will present examples of fragment-based approaches in lead discovery, categorized into the following four types: fragment evolution; fragment linking; fragment self-assembly and fragment optimization (BOX 2). TABLES 2–6 show the structures of the starting fragments and the optimized lead molecules from the 25 examples discussed.

Lead identification by fragment evolution

Identification of initial fragments using a direct binding technique is most useful if it is supported by some information about the binding mode of the fragment. With this type of information it is possible to develop hypotheses about how to build up larger and more complex molecules that target additional interactions in the active site of the protein. This 'evolution' leads to tighter-binding molecules, which can then be further optimized (FIG. 4; TABLE 2).

RULE OF FIVE
Identifies several key properties that should be considered for compounds with oral delivery in mind. These properties are molecular mass <500 Da, cLogP <5, number of hydrogen-bond donors ≤5 and number of hydrogen-bond acceptors ≤10.

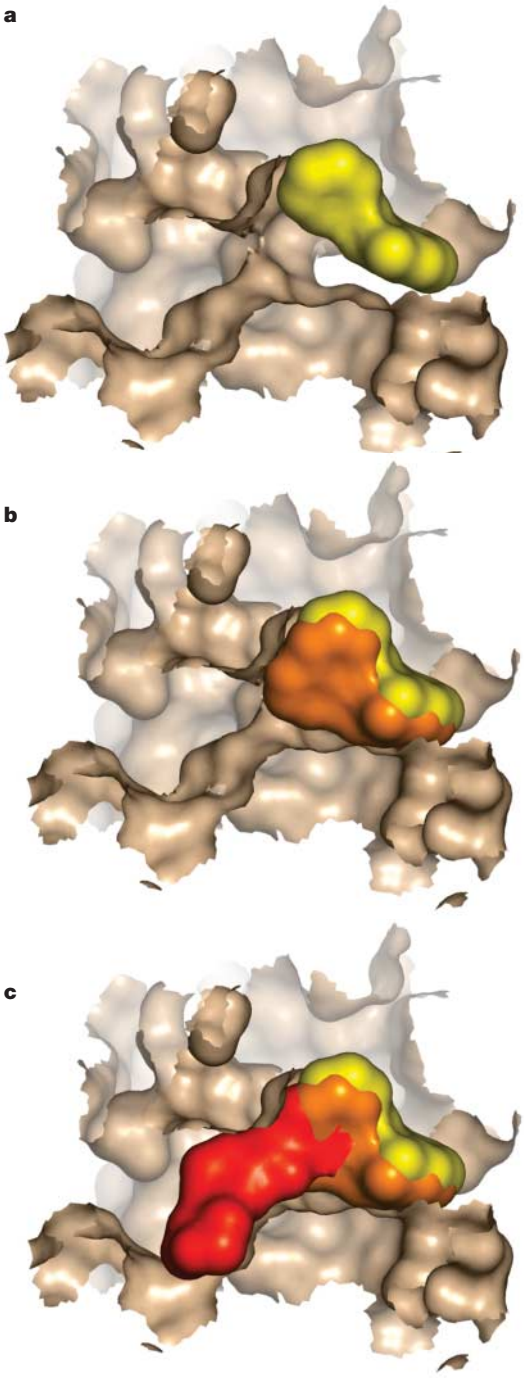


Figure 1 | **Surface representation of fragment growth against p38 mitogen-activated protein kinase. a** | The starting fragment is shown in yellow binding at the ATP-binding site (potency >1 mM, M_r = 200). **b** | A compound is displayed in orange (potency ~50 μM, M_r = 250) that was derived from the first chemistry iteration on the starting fragment (also shown in yellow). **c** | A subsequent lead molecule is shown in red (potency ~ 300 nM, M_r = 425) superimposed on the earlier fragments. All binding modes are taken from crystal structures and in all cases the displayed protein surface is taken from the crystal structure of the red molecule in **c**. This was done to illustrate the evolution of the fragment in a clearer way, because there are changes in protein conformation in going from **a** to **c**[53]. The protein surface has been clipped to allow a clearer view of the inhibitors. The figure is based on data obtained during an Astex Technology programme.

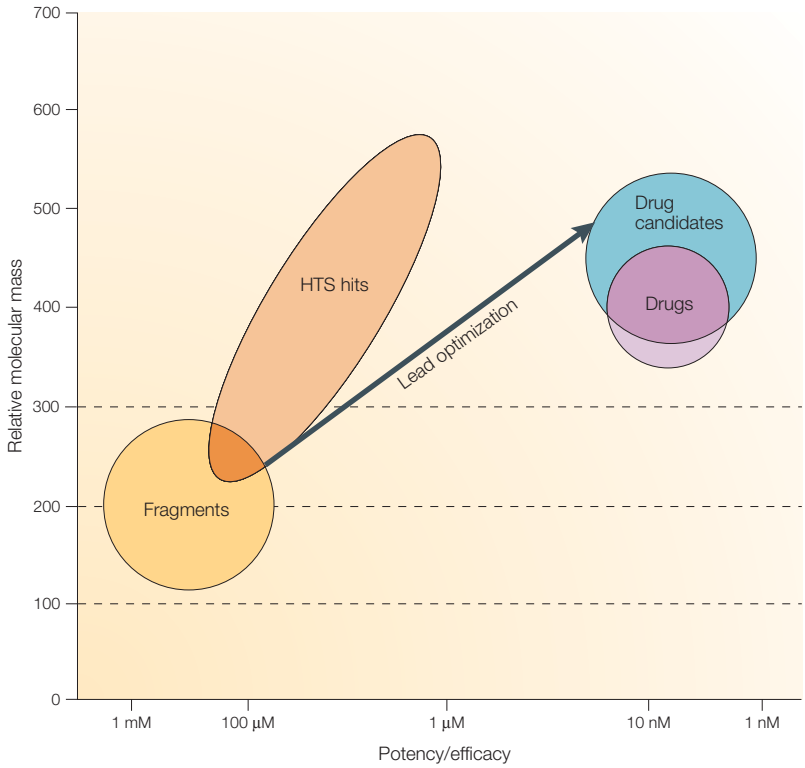


Figure 2 | **Schematic comparison of the usual molecular mass and potency ranges of high-throughput screening hits with fragments as starting points for lead identification and drug discovery.** The figure shows graphically a broad generalization of the range of molecular mass and potency for high-throughput screening (HTS) hits and fragments, superimposed on the typical requirements for leads, drug candidates and oral drugs using the same criteria. Fragment hits will have an M_r in the range 120–250 and low potency (mM–30 μM); to become useful leads, fragment potency will need to be increased, almost always with a resulting increase in M_r. HTS hits will have a much broader range of M_r (perhaps 250–600) and tend to be in the low-μM to high-nM potency range. Often, M_r will need to be reduced and potency retained or increased to produce a quality lead series. Lead molecules themselves tend to be relatively potent for their size — for example, 100-nM potency for M_r = 350 (REF. 13). Oral small-molecule drug candidates tend to have M_r <500 and be highly potent and efficacious, and launched oral drugs tend to have M_r significantly below 500[14,15]. There are, of course, many exceptions to the above historical generalizations.

Table 2, entry 1. A method referred to as 'needle screening' has been used to identify inhibitors that bind to the ATP-binding site of the bacterial enzyme DNA gyrase[1]. Fourteen classes of needle hits (or fragments) were identified by *in vitro* bioassay and validated by biophysical methods including NMR, SURFACE PLASMON RESONANCE (SPR) and X-ray crystallography. Subsequent three-dimensional-structure guided optimization using information obtained from the X-ray crystal structure of the ATP-binding pocket led to a compound that is reported to be >10,000-fold more active than the starting indazole fragment (maximal non-effective concentration (MNEC) in DNA gyrase inhibition 0.03 and >250 μg per ml, respectively). The authors suggest that needle screening provides chemical starting points that have no unnecessary structural elements, and therefore reduces the risk of toxicity or metabolic instability.

Table 2, entry 2. A native or engineered cysteine in a protein was allowed to react with a small library of disulphide-containing molecules (~1,200 compounds) at concentrations of 10–200 μM under conditions that favour disulphide exchange[19]. The cysteine-captured ligands (not shown in the table) are readily identified by mass spectrometry and form stable complexes, even though in the absence of the tethering the ligands might bind very weakly. This approach was used to generate a very potent inhibitor of thymidylate synthase. The affinity of the untethered ligand (~mM) was improved 3,000-fold by the synthesis of a small set of analogues guided by crystallographic structures of the tethered ligands.

Table 2, entry 3. Fragment chemistry was used to track the activity of inhibitors of the ATP-binding site of p38 mitogen-activated protein (MAP) kinase[2]. On the basis of a library of carefully selected drug-like scaffolds, and using NMR-based screening, a fragment with K_d = 1 mM was identified; by adding first one and then a second aromatic ring on to the central five-membered heterocyclic ring, activity reached K_i = 200 nM.

Table 2, entry 4. High-throughput X-ray crystallography screening of fragments using crystals of p38 MAP kinase led to the identification of a pyridyl indole derivative that has IC_{50} = 33 μM[20]. This compound was one of a range of different hits identified by interpretation of electron density in the active site following X-ray structure determination of fragment-soaked crystals. Fragment evolution based on the X-ray structure led to the lead illustrated (IC_{50} = 142 nM), and required the synthesis of only 70 compounds.

Table 2, entry 5. Fragment evolution against urokinase was initiated from the benzamidine analogue naphthamidine, and resulted in a 1,000-fold increase in activity[21]. The optimization was driven by cycles of structure-based design (SBD), and significant increases in selectivity were also observed over the starting fragments. The results are impressive because outside of the S1 specificity pocket in which the amidines bind, urokinase exhibits a relatively flat solvent-exposed active site compared with other serine proteases, such as thrombin and factor Xa.

In another example (not shown in TABLE 2), a submicromolar inhibitor against thrombin was discovered starting from aminobenzamidine (K_i = 34 μM). Using structure-based drug design, ten compounds were synthesized, and one compound showed a ~1,000-fold improvement in potency (K_i = 95 nM)[22].

Table 2, entry 6. Structure-based screening was applied to block the binding of endogenous ligands to human adipocyte fatty-acid-binding protein FABP4 (REF. 23). Hits were initially located via an NMR screen and then subsequently crystallized in the protein. Limited structure-based optimization led to a low-molecular-mass 10 μM lead. Crystallization of this lead molecule showed that the binding mode for the initial fragment was retained in the more decorated molecule.

SURFACE PLASMON RESONANCE (SPR). A phenomenon which occurs when light is reflected off thin metal films to which target molecules are immobilized and addressed by ligands in a mobile phase. If binding occurs to the immobilized target then the local refractive index changes, which leads to the apparent rate constants for the association and dissociation phases of the reaction. The ratio of these values gives the apparent equilibrium constant (affinity).

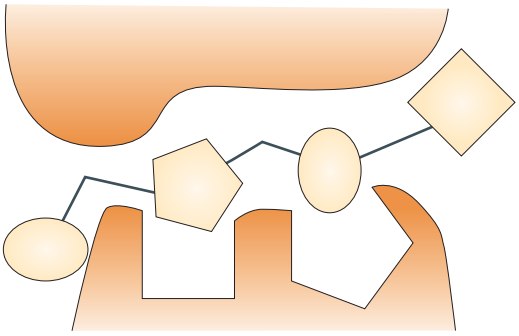


Figure 3 | **Schematic representation of a low-quality HTS hit.** The high-throughput screening (HTS) hit is large and makes surface contact with the receptor without forming high-quality interactions in key pockets. The affinity is spread throughout the entire molecule and, in the absence of structural information, the medicinal chemist does not know which areas of the molecule to focus on during hit optimization. Experience shows that optimization of these kinds of hits is very difficult. This is in contrast to the schematic in FIG. 4, in which fragment 1 is much smaller, makes high-quality contacts with the receptor and has relatively weak affinity. It has been shown that such fragments can often be built up into attractive leads with the aid of structural information (for example, TABLE 2)

Table 2, entry 7. NMR screening (that is, structure–activity relationships (SAR) by NMR[7]) was used to identify a series of triazine-containing compounds that bind in the mM range to ErmAM methyl transferase[24] (for example, TABLE 2, entry 7, NMR K_d = 1 mM), an enzyme target for ameliorating antibiotic resistance. Optimization of this initial lead using parallel synthesis led to inhibitors in the low μM range (for example, table entry K_i = 7.5 μM). NMR and X-ray structures show that these non-nucleoside compounds bind to the *S*-adenosylmethionine-binding site on the Erm protein and that there is scope for the incorporation of additional binding interactions.

Lead identification by fragment linking

TABLES 3,4 show examples in which two fragments have been identified that bind in separate binding sites that are close enough to each other to be chemically linked (FIG. 5). For this to be an efficient lead-identification approach, one needs to both identify the initial fragments and also have a process that allows the appropriate linking to be achieved in an efficient manner.

The potency increase achievable from optimally linking two fragments is often assumed to benefit from an approximate additivity effect, such that the free energy of binding of the joined molecule is approximately equal to the sum of the free energies of binding of the fragments (that is, two millimolar fragments when joined together lead to a micromolar inhibitor)[5]. Such additivity requires that the contribution from the linker is negligible and that the loss in rigid-body entropy on binding of all components to the enzyme is very small. Recently, an analysis of the experimental energetics associated with optimally linked fragments has suggested that the rigid-body entropy loss on protein binding constitutes a barrier of around three orders of magnitude to the binding affinity, and that this barrier is essentially independent of molecular mass[25]. The analysis implies that there should be a super-additivity effect when two fragments are linked in an optimal fashion. Such super-additivity is observed for entries 2 and 7 in TABLE 3.

Table 3, entry 1. SAR by NMR[7] was used to identify a potent 49 nM inhibitor of the FK506-binding protein (FKBP) binding domain by linking two weaker inhibitors (2 μM and 100 μM). NMR screening of a set of 1,000 fragments — including pipecolinic acid derivatives, a class of compounds known to bind to FKBP — identified the pipecolinic acid (K_d = 2 μM) and the diphenyl amide (K_d = 100 μM). Use of ^{15}N-^{13}C-filtered protein–ligand NUCLEAR OVERHAUSER EFFECT (NOE) data

Table 1 | **Comparison of fragment-based approaches and high-throughput screening**

Fragment-based approaches	High-throughput screening (HTS)
Emphasis on efficiency	Emphasis on potency
Typically screen a few hundred–few thousand compounds	Typically screen hundreds of thousands of compounds
M_r range ~150–300	M_r range ~250–600
Hit activity in the range mM–30 μM	Hit activity in the range ~30 μM–nM
Hits have clearly defined binding interactions; high proportion of atoms directly involved in protein binding	Hits can contain functional groups that contribute poorly to protein binding or act primarily as scaffolding (as shown schematically in FIG. 3)
Biophysical screening techniques (NMR, X-ray) are direct measurements of binding interaction. Can screen against 'inactive' forms of the target protein (for example, kinases)	*In vitro* bioassay-based screening. Can generate false positives and high attrition in hit-validation stage
Protein-structure-based information key in validating and prioritizing chemistry hits	Chemistry (re)synthesis resource usually required to validate and prioritize screening hits
Hit-to-lead chemistry usually requires synthesis of only a few compounds designed to add additional, specific binding interactions	High attrition of chemical series in hit-to-lead stage. Usually requires several iterations of high-throughput chemistry. Attrition rates can be improved with knowledge of protein structure
Design-intensive	Resource-intensive
Requires expertise and knowledge in protein structure, protein–ligand-binding interactions and fragment design	HTS requires extensive infrastructure for storing and handling compound collections, screening, automation, data processing and chemistry follow-up

NMR, nuclear magnetic resonance.

NUCLEAR OVERHAUSER EFFECTS (NOEs). Changes in the intensity of NMR signals, which are caused by through-space dipole–dipole coupling. Upper distance constraints obtained from ^{1}H–^{1}H NOEs are used for NMR structure determination of biological macromolecules.

Box 1 | Rule of three

The properties of 40 *fragment* hits identified against a range of targets using high-throughput X-ray crystallographic screening technology has been examined[18]. The results indicated that on average fragment hits possessed properties consistent with a 'rule of three' in which:

- **M_r < 300**
- **Number of hydrogen-bond donors ≤ 3**
- **Number of hydrogen-bond acceptors ≤ 3**
- **cLogP = 3**

In addition, it was noted that:

- **The number of rotatable bonds was, on average, ≤ 3**
- **Polar surface area was = 60 Å^2**

allowed sufficient information on the binding site to be obtained to guide the fragment linking.

Table 3, entry 2. Key questions concerning the efficiency of fragment linking have been addressed by detailed analysis of avidin binding by biotin analogues[26]. Femtomolar inhibitors were fragmented into their constituent parts and their affinities were measured. The example illustrates the super-additivity in binding affinity that can be achieved when two fragments are joined together in a near-optimal fashion.

Table 3, entry 3. The binding properties of VANCOMYCIN derivatives have been extensively studied with multivalent acetyl-Lys-D-Ala-D-Ala molecules[27,28]. Here, only the dimeric analogue is represented in the table. There is a large increase in affinity despite the introduction of additional rotatable bonds.

Table 3, entry 4. Starting from the crystal structure of an acridine analogue in acetylcholine esterase, it was discovered that a second molecule of the acridine analogue could be modelled in the active site[29]. Several dimeric analogues with different linker lengths were synthesized, and an inhibitor that was 1,000-fold more potent than the original molecule was identified.

Table 3, entry 5. The quaternary structure of human mast cell tryptase, a trypsin-like serine protease[30], has been used to motivate the design of linked dimers of benzamidine[31]. An aspect of this approach was the key observation that tryptase exists as a tetramer, and thereby provides the opportunity to couple ligands to form linked dimeric binders. The benzamidines illustrated in TABLE 3 were designed to bind into the S1 pockets of neighbouring tryptase monomers with a flexible linker region spanning the space between the two monomers, and the SAR of these inhibitors is supportive of this mechanism. It is interesting to note that APC-2059 (REF. 32), a tryptase inhibitor that has advanced to Phase II clinical trials, is also a dibasic inhibitor with an extended linking group (although this inhibitor was not derived directly from the fragment-linking approach).

Table 3, entry 6. A fragment library of oximes was screened at high concentration against the non-receptor tyrosine kinase c-SRC to identify two weak hits (IC_{50} = 40 μM and 41 μM, respectively)[3]. The two respective aldehydic precursors of the active oximes were then linked via a small library of di-hydroxylamine linkers to scope out and identify the optimum linking spacer group. This approach allowed the identification of a potent linked di-oxime (IC_{50} = 64 nM) without any structural knowledge of the target or binding sites.

Table 3, entry 7. An elegant approach to lead identification has been developed that uses NMR screening of a secondary site in the presence of a small ligand that binds to a known active-site 'hot spot'[33]. In this way, a simple bi-phenyl analogue that binds to the P1′ SITE of stromelysin with a K_d = 280 μM was identified in the presence of acetohydroxamic acid, which itself binds to the zinc ion in the catalytic site. Subsequently, an NMR structure of two fragments bound in the active site of the protein was solved. This allowed structure-guided design of a linked compound, which was found to be a very potent lead (K_d = 15 nM).

Table 4, entry 1. A technology termed extended tethering, related to that described in TABLE 2 (entry 2), was used to identify a novel non-peptide inhibitor of

Box 2 | Summary of fragment-based approaches

Fragment evolution
An initial fragment is optimized by adding functionality to bind to adjacent regions of the active site (illustrated schematically in FIG. 4, with examples given in TABLE 2).

Fragment linking
Two (or more) fragments, which bind to proximal parts of the active site, are joined together to give a larger, higher-affinity-binding molecule (illustrated schematically in FIG. 5, with examples given in TABLES 3,4).

Fragment self-assembly
Fragments with complementary functional groups are allowed to react together in the presence of the protein target and the most potent larger molecule is detected (illustrated schematically in FIG. 6 with examples given in TABLE 5). This includes approaches usually termed dynamic combinatorial chemistry.

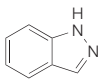
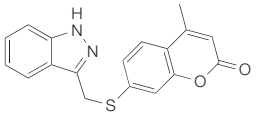
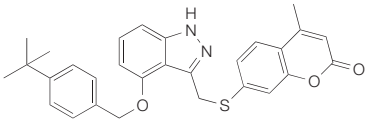
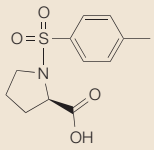
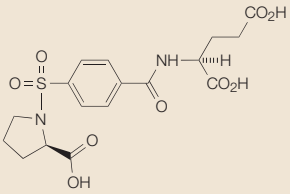
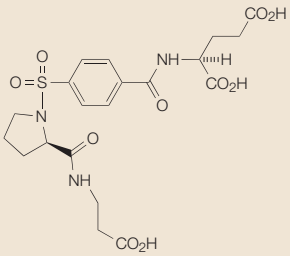
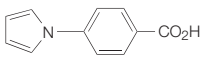
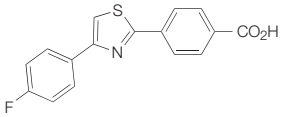
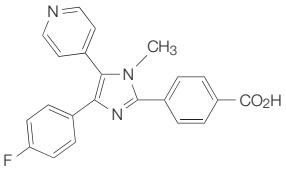
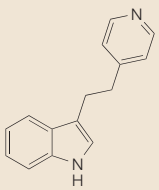
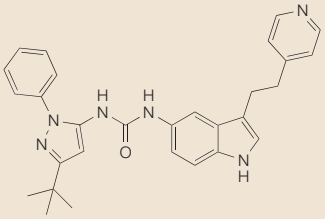
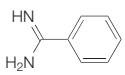
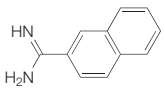
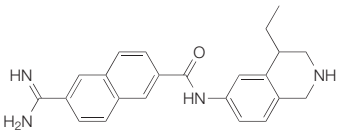
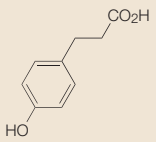
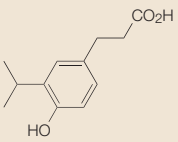
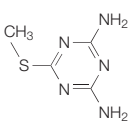
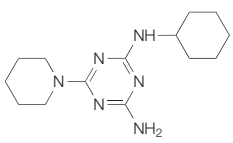
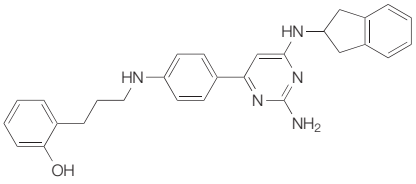
Fragment optimization
Fragment approaches are used to optimize drug-like properties of a lead other than just binding affinity (illustrated schematically in FIG. 7, with examples given in TABLE 6).

VANCOMYCIN
Vancomycin is an antibiotic that acts by binding to cell-wall precursors that terminate in the sequence D-Ala-D-Ala, thereby inhibiting cell-wall synthesis.

P1′ SITE
The substrate residue that occurs immediately after the scissile amide bond in a protease. It is the key specificity site of matrix metalloproteases like stromelysin.

caspase-3, a cysteine protease[34]. Mass spectrometry was used to identify the library members that had formed a disulphide bond to the tethered thiol. Subsequently, these library fragments were coupled to known reversible cysteine-binding elements to generate potent reversible molecules.

Table 2 | **Lead identification by fragment evolution***

Entry	Target/method	Fragment	Evolved fragment	Lead
1	DNA gyrase[1]/ VS and SBD			
		K_d = 10 mM (by NMR) MNEC >250 μg per ml	MNEC = 8 μg per ml	MNEC = 30 ng per ml
2	Thymidylate synthase[19]/ tethering and SBD			
		IC_{50} = 1.1 mM	IC_{50} = 24 μM	IC_{50} = 330 nM
3	p38 kinase[2]/ NMR			
		K_d = 1 mM	K_d = 200 μM	K_i = 200 nM
4	p38 kinase[20]/ X-ray and SBD			
		IC_{50} = 33 μM		IC_{50} = 142 nM
5	Urokinase[21]/ bioassay and SBD			
		K_i = not reported	K_i = 5.9 μM	K_i = 6.3 nM
6	FABP4 (REF. 23)/ NMR and SBD			
		IC_{50} = 590 μM		IC_{50} = 10 μM
7	Erm methyl transferase[24]/ NMR			
		K_d = 1 mM (by NMR)	K_d = 75 μM	K_i = 7.5 μM

*See also FIG. 4. FABP, fatty-acid-binding protein; MNEC, maximal non-effective concentration; NMR, nuclear magnetic resonance; SBD, structure-based design; VS, virtual screening.

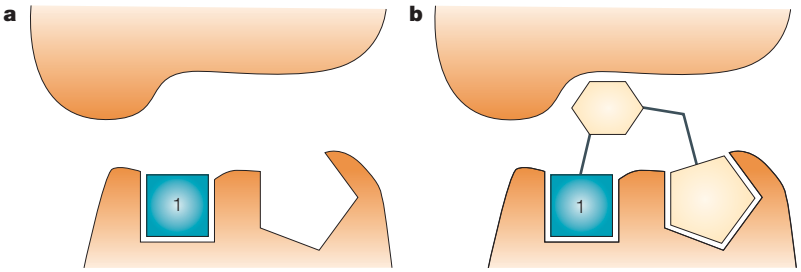


Figure 4 | **Fragment evolution. a** | Fragment 1 binds to the receptor at one site. **b** | The lead molecule is evolved by building away from the starting fragment and making good contact with the upper surface and then by growing into a second pocket. For examples, see TABLE 2.

Table 4, entries 2 and 3. Fragment linking has been applied to the extremely challenging target protein tyrosine phosphatase-1B[35,36]. Phosphatases dephosphorylate peptide substrates by recognizing the doubly acidic phosphate group and key peptide residues; it is therefore not surprising that it is difficult to identify drug-like small-molecule leads against phosphatases. NMR screening of the catalytic phosphate-binding site, and a secondary phosphate-binding site, enabled the identification of a potent tool molecule (entry 2) and a useful hit (entry 3) by fragment linking.

Table 4, entry 4. Mass spectrometry has been used as the method for detecting the initial fragment hits for the 1061 region of bacterial 23S rRNA, a technique referred to as 'SAR by MS'[37]. This subdomain of the ribonucleic acid is part of the binding site for the antibiotic thiostrepton. *In vitro* binding experiments showed that the two fragments bind to different sites on the RNA; subsequently, several fused compounds were synthesized that have markedly tighter binding. The authors reported that traditional HTS assays for this antibacterial target gave very low hit rates.

Table 4, entry 5. A potent small-molecule inhibitor of interleukin-2 (IL-2) was identified through the use of fragment tethering and fragment assembly[38]. Analysis of the X-ray structure of a known 3-μM inhibitor revealed that the protein is adaptive and able to undergo significant rearrangement, which creates small-molecule-binding sites. Ten individual cysteine mutants were designed to search the perimeter of the IL-2 binding 'hot spot'. These mutants were then screened against a library of 7,000 disulphide-containing fragments. Analysis of the bound disulphide-tethered fragments indicated that these fragments could occupy a deep hydrophobic cavity within the adaptive region. An overlay of the modelled tethered hits with a known crystal structure indicated a coupling of two of the fragments and allowed identification of sub-100 nM inhibitors.

Lead identification by fragment self-assembly

The use of reactive fragments that are capable of self-assembly in the presence of a template molecule (such as a protein) (FIG. 6) is a large and growing field[39]. TABLE 5 cites some key examples in which two separate fragments are linked together to form a larger and more active inhibitor in the presence of the protein target itself.

These examples are categorized here as 'fragment self-assembly' because the protein is used to self-select or to catalyse the synthesis of its own inhibitor without covalent attachment of the protein to the inhibitor. This definition forms the basis for distinguishing the examples in TABLE 6 from those in TABLE 5.

Table 5, entry 1. In this prototype example, a mixture of imines was prepared from four amines and three aldehydes under reversible conditions and then reduced to the corresponding amines[40]. High-performance liquid chromatography was used to demonstrate the presence of all possible amines. When the same reaction is performed in the presence of carbonic anhydrase, the proportion of one amine is increased and this is presumed to correspond to the strongest inhibitor. This is referred to as dynamic combinatorial chemistry[39].

Table 5, entry 2. Similar dynamic combinatorial chemistry with imines has been used to identify a neuraminidase inhibitor from diamine and ketone building blocks[41,42]. The amplification factor in the formation of the secondary amine (table entry K_i = 85 nM), as determined by liquid chromatography–mass spectrometry (LC–MS), was >30. The structure of this lead is closely related to the active component of oseltamivir (Tamiflu; Roche), a marketed influenza neuraminidase inhibitor.

Table 5, entry 3. A complementary approach has been reported in which inhibitors are formed from a dynamic combinatorial library in the presence of protein crystals and the binary complex is then observed directly by X-ray crystallography[43]. This has been termed 'dynamic combinatorial X-ray crystallography' (DCX) and is illustrated by the identification of a previously reported inhibitor of cyclin-dependent kinase-2 (table entry IC_{50} = 30 nM) from a mixture of essentially inactive hydrazines and isatins as potential fragments for adjacent binding pockets within the ATP site.

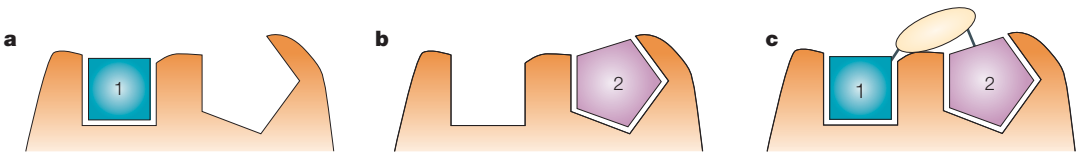
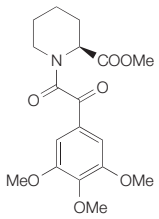
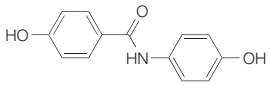
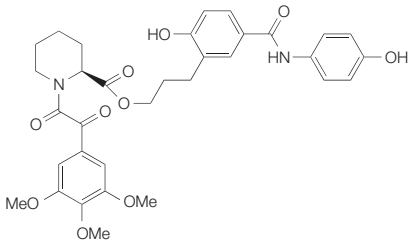
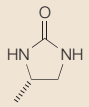
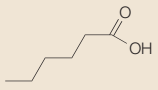
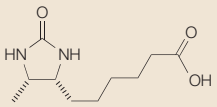
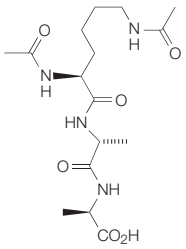
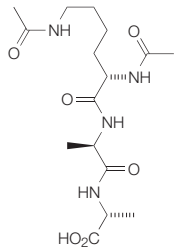
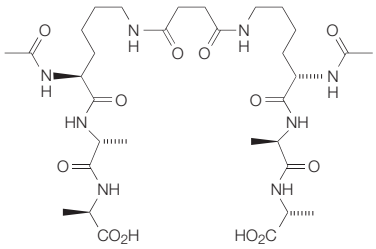
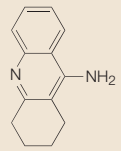
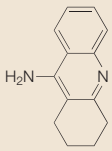
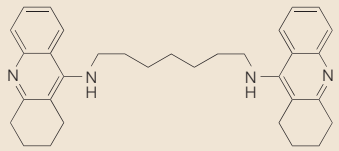
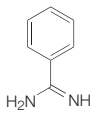
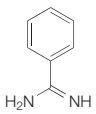
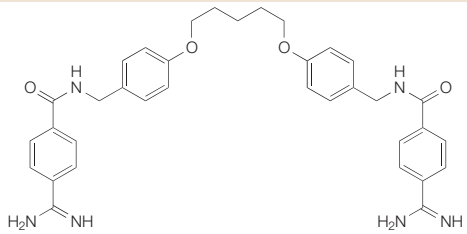
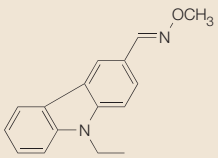
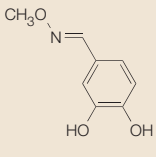
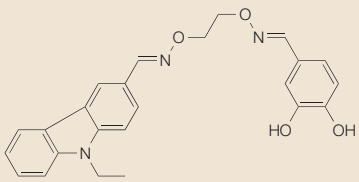
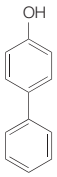
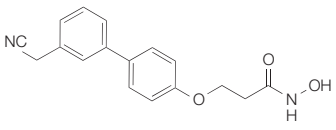


Figure 5 | **Fragment linking. a** | Fragment 1 binds to the receptor at one site. **b** | Fragment 2 binds to the receptor at an adjacent site. **c** | Fragments joined together by a linking group that allows the lead molecule to span both sites. For examples, see TABLES 3,4.

Table 3 | **Lead identification by fragment linking: part 1***

Entry	Target/method	Fragments		Lead
1	FKBP[7]/ NMR			
		K_d = 2 μM	K_d = 100 μM	K_d = 49 nM
2	Avidin[26]/ bioassay			
		–ΔG = 6.1 Kcal per mol (K_i = 34 μM)	–ΔG = 4.9 Kcal per mol (K_i = 260 μM)	–ΔG = 16.9 Kcal per mol (K_i = 0.0004 nM)
3	Vancomycin[27,28]/ bioassay			
		K_i = 4.8 μM	K_i = 4.8 μM	K_i = 1.1 nM
4	Acetylcholine esterase[29]/ bioassay and SBD			
		K_i = 0.6 μM		K_i = 0.4 nM
5	Tryptase[31]/ bioassay and SBD			
		K_i = 22 μM	K_i = 22 μM	K_i <0.01 nM
6	c-SRC[3]/ bioassay			
		IC_{50} = 40 μM	IC_{50} = 41 μM	IC_{50} = 64 nM
7	Stromelysin[33]/ NMR and SBD			
		K_d = 17 mM	K_d = 280 μM	K_d = 15 nM

*See also FIG. 5. FKBP, FK506-binding protein; NMR, nuclear magnetic resonance; SBD, structure-based design

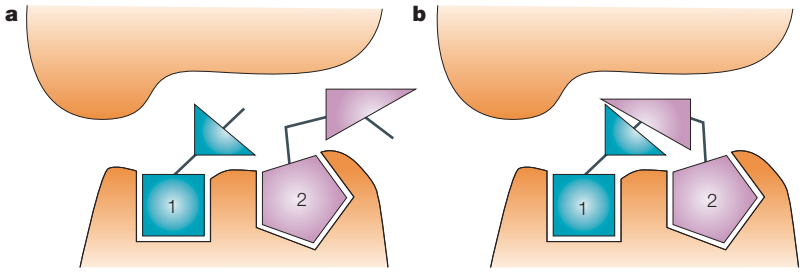


Figure 6 | **Fragment self-assembly. a** | Fragments 1 and 2 bind to receptor sites simultaneously with reacting groups positioned within conformational reach of each other, increasing the effective molarity of reacting groups. **b** | Lead molecule formed in the active site. For examples, see TABLE 5.

Table 5, entry 4. The building blocks are substantially larger and higher affinity than many of the fragments described above, but the example is included to illustrate the concept of self-assembly[44,45]. A femtomolar inhibitor of acetylcholinesterase (K_d = 77 fM) has been identified from an array of building blocks containing an acetylene or azide group which undergo an irreversible 1,3 dipolar cyclo-addition reaction as a result of binding to the enzyme. This is referred to as 'click chemistry *in situ*'. The enzyme binds both fragments simultaneously, thereby positioning the azide and acetylene groups sufficiently close to each other to allow them to react together much faster than would otherwise be the case. Structural knowledge of the binding interactions of acetylcholinesterase inhibitors facilitated the selection of these particular fragments.

Lead progression by fragment optimization

TABLE 6 shows examples in which the fragment approach has been used to optimize or modify properties other than just binding potency of a lead (FIG. 7).

Table 6, entry 1. This example illustrates the identification of a clinical candidate from a fragment-based discovery programme[4]. The known binder, benzamidine (200 μM), was chosen as a seed template for the S1 POCKET of factor Xa. A structure-based virtual screening method was used to target proximal enzyme pockets and to drive three iterations of chemical synthesis. This resulted in the compound shown in TABLE 6, which has an IC_{50} = 16nM and is more than 10,000 times more potent than the initial fragment. Subsequent optimization using a combination of medicinal chemistry and structure-based drug design led to the replacement of the benzamidine, a moiety that is often associated with poor oral bioavailability. Further lead optimization resulted in the discovery of the orally bioavailable candidate LY517717, and at a recent conference presentation it was disclosed that this compound has successfully completed Phase I clinical trials, with clinical development ongoing. This example demonstrates that it is possible to generate an advanced drug candidate starting from a fragment with weak potency.

S1 POCKET
The pocket on a protease occupied by the substrate residue which immediately precedes the scissile amide bond. It is the key specificity pocket of trypsin-like serine proteases such as factor Xa, urokinase, tryptase and thrombin, in which lysine or arginine are the favoured substrate residues.

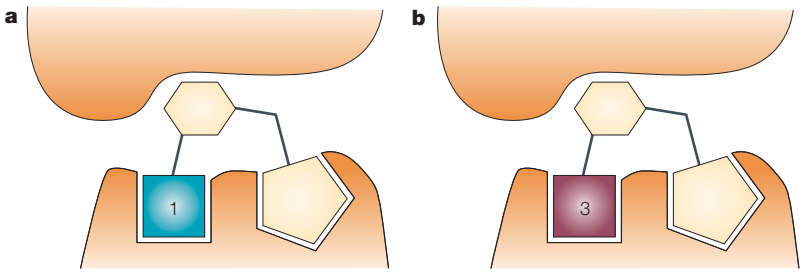


Figure 7 | **Lead progression via fragment optimization. a** | Existing lead molecule discovered by fragment-based approach. **b** | Lead molecule re-engineered to address optimization of a particular property (for example, selectivity, cell-based activity, oral activity or efficacy). For examples, see TABLE 6.

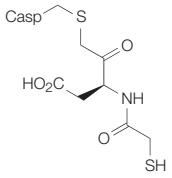
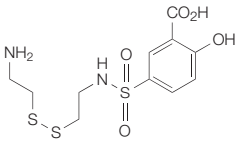
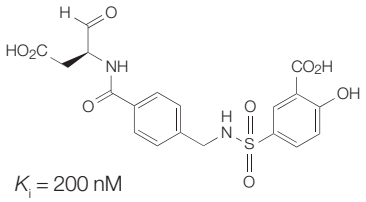
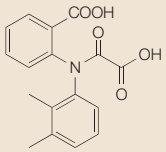
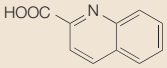
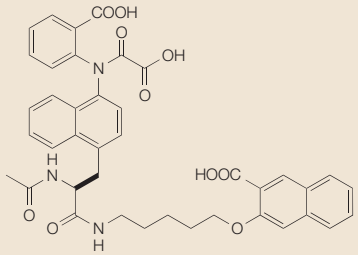
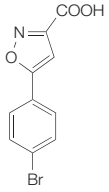
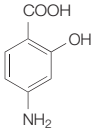
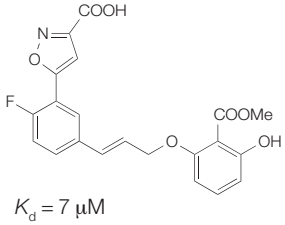
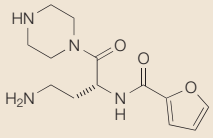
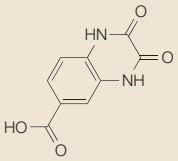
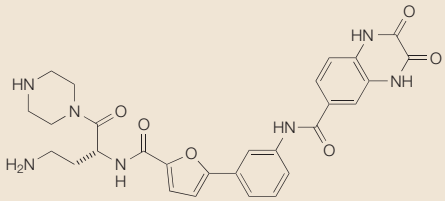
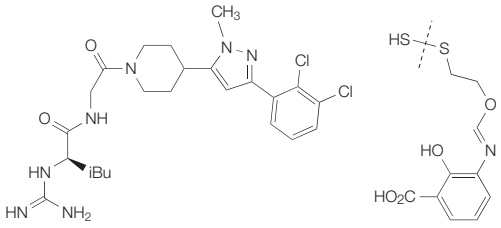
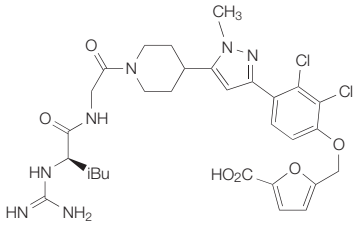
Table 6, entry 2. One of the challenges in discovering useful urokinase inhibitors is to identify leads that do not have highly basic amidine or guanidine groups (pK_a >9). An NMR screen of more than 3,000 compounds led to the identification of 2-aminobenzimidazole as a weak (IC_{50} = 200 μM) but competitive inhibitor that binds to the same site on urokinase as the more traditional inhibitors, but is less ionized at physiological pH (pK_a = 7.5)[46]. Screening simple analogues of this hit led to the 5-hydroxy derivative, IC_{50} = 10 μM (pK_a = 7.4), and an X-ray co-complex structure was determined which could be used for subsequent optimization.

Table 6, entry 3. The same target, urokinase, has been the subject of a fragment screen using X-ray crystallography[47]. Mixtures of weakly basic compounds were assessed as urokinase ligands by crystal-soaking experiments. From a total of some 61 compounds, 8-hydroxy-2-amino quinoline was identified, and had K_i = 56 μM and pK_a = 7.3. An attraction of the X-ray screening technique is that it generates structural information that can be used directly in subsequent lead optimization. In this case, the optimized compound had a 100-fold increase in enzyme activity (K_i = 370 nM), and, importantly, had 38% oral bioavailability.

Table 6, entry 4. Several drug discovery laboratories have targeted the sequence homology-2 (SH2) domain of the SRC protein family, and although nanomolar inhibitors have been identified, they are characterized by the presence of a phosphate group that is a liability in terms of its rapid hydrolysis and its contribution to poor cell penetration. An 'SAR by X-ray' approach has been developed and applied to find replacements for the phosphate group[48]. Small aromatic compounds were screened as phenyl phosphate surrogates by performing crystal-soaking X-ray structure determination for ~20 compounds. Optimization led to a phenyl tricarboxylic acid replacement for the phenyl phosphate that retained activity (IC_{50} = 3 nM) and was stable in rat plasma for more than 24 hours (versus the phosphate analogue $t_{1/2}$ = 0.6 hours). In a subsequent publication from the same laboratory[49], an SPR (Biacore) assay was used as a pre-screen before X-ray soaking of around 200 fragments to identify phosphotyrosine replacements.

Table 6, entry 5. A strategy to break down an existing chemical lead into its fragments and scaffold, identify replacements for the fragments and then incorporate

Table 4 | **Lead identification by fragment linking: part 2***

Entry	Target/method	Fragments		Lead
1	Caspase[34]/ tethering and SBD			 K_i = 200 nM
2	PTP1B[35]/ NMR and SBD	 K_d = 100 μM	 K_d >1,000 μM	 K_d = 22 nM
3	PTP1B[36]/ NMR and SBD	 K_d = 800 μM	 K_d = 1.2 mM	 K_d = 7 μM
4	Bacterial 23S rRNA[37]/ MS 'SAR by MS'	 K_d >100 μM	 K_d >100 μM	 K_d = 6.5 μM
5	IL-2 (REF. 38)/ Tethering and SBD	 IC_{50} = 3 μM		 IC_{50} = 60 nM

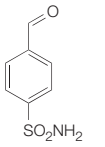
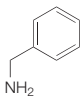
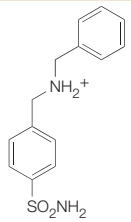
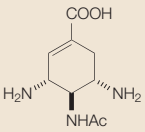
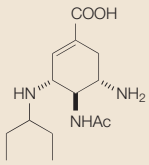
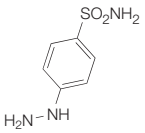
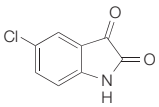
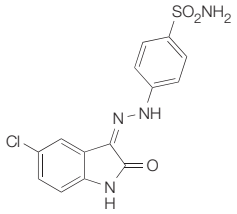
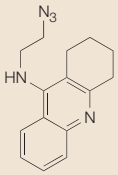
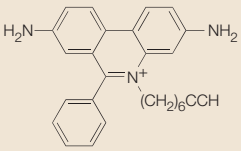
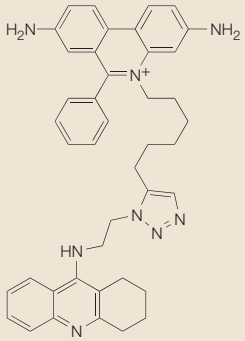
*See also FIG. 5. Casp, caspase; IL, interleukin; MS, mass spectrometry; NMR, nuclear magnetic resonance; PTP, protein tyrosine phosphatase; SAR, structure–activity relationship; SBD, structure-based design.

the newly identified fragments back into the original scaffold has been disclosed[50]. This offers the potential to modify existing leads in terms of pharmacokinetic properties or side-effect profile. A known adenosine kinase inhibitor (K_i = 1.7 nM) was fragmented, and NMR screening was used to identify alternative groups for a *meta*-bromo phenyl side chain fragment. An example was the indole ring system (K_d = 3 mM), and when this was incorporated into the ATP-hinge-binding heterocycle scaffold, it gave a modified lead (IC_{50} = 10 nM) that retained potency *in vitro* and *in vivo* in a hyperalgesia test.

Summary and outlook

The fragments described in this review have M_r = 120–250 and binding affinity in the range mM–30 μM. The weak absolute potency of fragments belies their high efficiency as ligands, because fragments are extremely potent for their size[51]. Furthermore, crystal structures of nanomolar leads that have been developed starting from fragments show very similar binding modes to those seen in the crystal structure of the isolated fragments, illustrating that fragments form very strong interactions with their proteins and provide good starting points for chemical optimization.

Table 5 | **Lead identification by fragment self-asembly***

Entry	Target/method	Fragments		Lead
1	Carbonic anhydrase[40]/ virtual combinatorial chemistry			
		K_i = not reported		
2	Neuraminidase[41,42]/ dynamic combinatorial libraries			
		K_i = not reported		K_i = 85 nM
3	CDK2 (REF. 43)/ DCX			
		IC_{50} >1 mM	IC_{50} >1 mM	IC_{50} = 30 nM
4	Acetylcholine esterase[45]/ click chemistry *in situ*			
		K_d = 10–100 nM	K_d = 10–100 μM	K_d = 77 fM

*See also FIG. 6. CDK, cyclin-dependent kinase; DCX, dynamic combinatorial X-ray crystallography.

Generally, fragments are identified using a biophysical screening method, most commonly NMR or protein crystallography, supported by a conventional enzyme bioassay. Consequently, information about the structure of the fragment–protein-binding interaction is generated as part of the screening. This structural information means that it is possible to incorporate a large element of design in optimizing the fragment into a high-affinity lead, either by growing additional binding groups or joining two fragments together. As a result, fragments can be optimized into nanomolar leads via the synthesis of significantly fewer compounds than in traditional approaches.

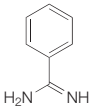
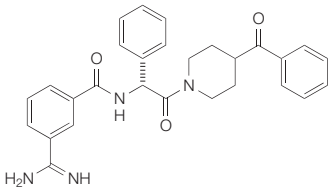
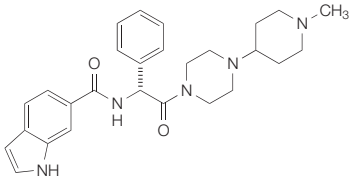
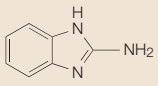
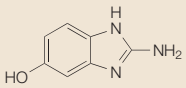
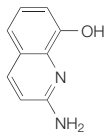
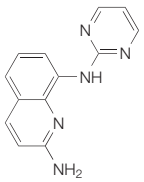
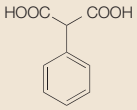
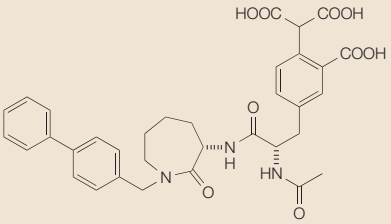
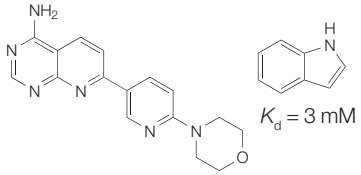
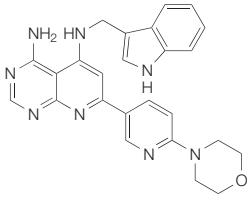
The examples in this review are derived more or less equally in a three-way split from academic, biotechnology and established pharmaceutical laboratories, and cover 25 protein targets, the majority of which are enzymes. In all cases, the affinity of the starting fragments is increased by orders of magnitude, and in ~80% of the examples the activity is driven to the nanomolar affinity level. Some of the protein targets are ones that have been regarded as difficult to obtain attractive small-molecule leads against using conventional screening. Although fragment-based lead discovery is still too young for its output to be measured in terms of clinical impact, one example of a compound in Phase II is cited in which the initial chemical approach used fragments (see TABLE 6, entry 1).

The technique of fragment-based lead discovery can be seen as orthogonal to that of HTS; the screening techniques, affinity, M_r, size of compound libraries and chemical strategies for the subsequent 'hit-to-lead' stage are all different. It therefore represents an alternative and complementary strategy, and also presents interesting challenges, both organizationally and culturally, for companies that already have significant infrastructure for HTS. A simplistic summary is that the fragment approach emphasizes efficiency and design, whereas HTS emphasizes affinity and numbers.

Furthermore, screening for low-M_r fragment hits, as expected, has a substantially higher hit rate than HTS. This efficiency is carried into the chemical optimization stage, as knowledge about the protein–fragment-binding interaction allows higher affinity to be achieved with only a few compounds being synthesized.

Proponents of the fragments approach (see REF. 52 for a recent complementary review) argue that starting with a low-M_r compound in which the entire molecule is needed for the target protein binding increases the chance of discovering molecules with the desired physicochemical properties of oral drugs (FIG. 2). In view of the widespread recognition that lead optimization is usually accompanied by increasing M_r by 100–150, and that increasing M_r seems to correlate with attrition in clinical development, it is particularly attractive to start lead optimization with M_r <300. However, in the past, hits have often been selected on the basis of criteria associated with absolute potency or binding affinity, whereas the fragment approach emphasizes binding efficiency; that is, binding energy divided by M_r. Distinctive chemical strategies have been devised for building low-M_r fragments into nanomolar leads. Fragment evolution involves growing out from an initial fragment, whereas in the fragment-linking strategy two separate fragments are linked together. In these respects, the fragment approach represents a new way of working for hit and lead identification in drug discovery projects.

Table 6 | **Lead progression by fragment optimization***

Entry	Target/method	Fragments	Lead
1	Factor Xa[4]/ bioassay and SBD	 K_i = 200 μM	 K_i = 16 nM  Subsequently optimized compound, LY517717, currently in Phase II
2	Urokinase[46]/ NMR	 K_d = 200 μM	 K_d = 10 μM pK_a = 7.4
3	Urokinase[47]/ X-ray screen and SBD	 K_i = 56 μM	 K_i = 370 nM 38% oral bioavailability
4	SH2 domain of pp60Src[48,49]/ X-ray screen, SPR and SBD	 IC_{50} = 2.5 mM	 IC_{50} = 3 nM Stable in rat and human plasma >24 hours
5	Adenosine kinase[50]/ NMR	 K_d = 3 mM	 IC_{50} = 10 nM *In vivo* hyperalgesia data

*See also FIG. 7. NMR, nuclear magnetic resonance; SBD, structure-based design; SH2, sequence homology-2; SPR, surface plasmon resonance.

1. Boehm, H. J. *et al*. 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**, 2664–2674 (2000).
An early description of a strategy to identify needles via virtual screening, *in vitro* screening and validation by biophysical methods; also provides an example of fragment evolution applied to DNA gyrase.
2. Fejzo, J. *et al*. The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* **6**, 755–769 (1999).
3. Maly, D. J., Choong, I. C. & Ellman, J. A. Combinatorial target-guided ligand assembly: identification of potent subtype-selective c-Src inhibitors. *Proc. Natl Acad. Sci. USA* **97**, 2419–2424 (2000).
4. Liebeschuetz, J. W. *et al*. PRO_SELECT: combining structure-based drug design and array-based chemistry for rapid lead discovery. 2. The development of a series of highly potent and selective factor Xa inhibitors. *J. Med. Chem.* **45**, 1221–1232 (2002).
5. Jencks, W. P. On the attribution and additivity of binding energies. *Proc. Natl Acad. Sci. USA* **78**, 4046–4050 (1981).
6. Farmer, P. S. & Ariens, E. J. Speculations on the design of nonpeptidic peptidomimetics. *Trends Pharmacol. Sci.* **3**, 362–365 (1982).
7. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531–1534 (1996).
Prototype description of the SAR by NMR technique and an early example of fragment linking to obtain nM ligands for the FK506-binding protein.
8. Pellecchia, M., Sem, D. S. & Wuthrich,K. NMR in drug discovery. *Nature Rev. Drug Discov.* **1**, 211–219 (2002).
9. Kuhn, P., Wilson, K., Patch, M. G. & Stevens, R. C. The genesis of high-throughput structure-based drug discovery using protein crystallography. *Curr. Opin. Chem. Biol.* **6**, 704–710 (2002).
10. Blundell, T. L., Jhoti, H. & Abell, C. High-throughput crystallography for lead discovery in drug design. *Nature Rev. Drug Discov.* **1**, 45–54 (2002).
Review of the techniques required for high-throughput X-ray crystallography and discussion of this as a screening technique for lead discovery.
11. Hann, M. M., Leach, A. R. & Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **41**, 856–864 (2001).
12. Oprea, T. I., Davis, A. M., Teague, S. J. & Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* **41**, 1308–1315 (2001).
13. Teague, S. J., Davis, A. M., Leeson, P. D. & Oprea, T. The design of leadlike combinatorial libraries. *Angew. Chem. Int. Ed. Engl.* **38**, 3743–3748 (1999).
14. Wenlock, M. C., Austin, R. P., Barton, P., Davis, A. M. & Leeson, P. D. A comparison of physiochemical property profiles of development and marketed oral drugs. *J. Med. Chem.* **46**, 1250–1256 (2003).
15. Vieth, M. *et al*. Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* **47**, 224–232 (2004).
16. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
17. Veber, D. F. *et al*. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).
18. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* **8**, 876–877 (2003).
19. Erlanson, D. A. *et al*. Site-directed ligand discovery. *Proc. Natl Acad. Sci. USA* **97**, 9367–9372 (2000).
20. Frederickson, M., Gill, A. L., Padova, A. & Congreve, M. S. Preparation of indoles as p38 MAP kinase inhibitors. (Astex Technology Limited, UK. UK Patent WO 03/087087 (2003).
21. Wendt, M. D. *et al*. Identification of novel binding interactions in the development of potent, selective 2-naphthamidine inhibitors of urokinase: synthesis, structural analysis, and sar of *N*-phenyl amide 6-substitution. *J. Med. Chem.* **47**, 303–324 (2004).
22. Bohm, H. J., Banner, D. W. & Weber, L. Combinatorial docking and combinatorial chemistry: design of potent non-peptide thrombin inhibitors. *J. Comput. Aided Mol. Des.* **13**, 51–56 (1999).
23. van Dongen, M. J. P. *et al*. Structure-based screening as applied to human FABP4: a highly efficient alternative to HTS for hit generation. *J. Am. Chem. Soc.* **124**, 11874–11880 (2002).
24. Hajduk, P. J. *et al*. Novel inhibitors of Erm methyltransferases from NMR and parallel synthesis. *J. Med. Chem.* **42**, 3852–3859 (1999).
25. Murray, C. W. & Verdonk, M. L. The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *J. Comput. Aided Mol. Des.* **16**, 741–753 (2002).
26. Green, N. M. Avidin. *Adv. Protein Chem.* **29**, 85–133 (1975).
27. Rao, J. & Whitesides, G. M. Tight binding of a dimeric derivative of vancomycin with dimeric L-Lys-D-Ala-D-Ala. *J. Am. Chem. Soc.* **119**, 10286–10290 (1997).
28. Rao, J., Lahiri, J., Weis, R. M. & Whitesides, G. M. Design, synthesis, and characterization of a high-affinity trivalent system derived from vancomycin and L-Lys-D-Ala-D-Ala. *J. Am. Chem. Soc.* **122**, 2698–2710 (2000).
29. Pang, Y. P., Quiram, P., Jelacic, T., Hong, F. & Brimijoin, S. Highly potent, selective, and low cost bis-tetrahydroaminacrine inhibitors of acetylcholinesterase. Steps toward novel drugs for treating Alzheimer's disease. *J. Biol. Chem.* **271**, 23646–23649 (1996).
30. Pereira, P. J. *et al*. Human β-tryptase is a ring-like tetramer with active sites facing a central pore. *Nature* **392**, 306–311 (1998).
31. Burgess, L. E. *et al*. Potent selective nonpeptidic inhibitors of human lung tryptase. *Proc. Natl Acad. Sci. USA* **96**, 8348–8352 (1999).
32. Rice, K. D. *et al*. Dibasic inhibitors of human mast cell tryptase. Part 2: structure–activity relationships and requirements for potent activity. *Bioorg. Med. Chem. Lett.* **10**, 2361–2366 (2000).
33. Hajduk, P. J. *et al*. Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.* **119**, 5818–5827 (1997).
34. Erlanson, D. A. *et al*. *In situ* assembly of enzyme inhibitors using extended tethering. *Nature Biotechnol.* **21**, 308–314 (2003).
35. Szczepankiewicz, B. G. *et al*. Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *J. Am. Chem. Soc.* **125**, 4087–4096 (2003).
36. Liu, G. *et al*. Fragment screening and assembly: a highly efficient approach to a selective and cell active protein tyrosine phosphatase 1B inhibitor. *J. Med. Chem.* **46**, 4232–4235 (2003).
37. Swayze, E. E. *et al*. SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *J. Med. Chem.* **45**, 3816–3819 (2002).
38. Braisted, A. C. *et al*. Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. *J. Am. Chem. Soc.* **125**, 3714–3715 (2003).
39. Ramstrom, O. & Lehn, J. M. Drug discovery by dynamic combinatorial libraries. *Nature Rev. Drug Discov.* **1**, 26–36 (2002).
40. Huc, I. & Lehn, J. M. Virtual combinatorial libraries: dynamic generation of molecular and supramolecular diversity by self-assembly. *Proc. Natl Acad. Sci. USA* **94**, 2106–2110 (1997).
41. Hochgurtel, M. *et al*. Ketones as building blocks for dynamic combinatorial libraries: highly active neuraminidase inhibitors generated via selection pressure of the biological target. *J. Med. Chem.* **46**, 356–358 (2003).
42. Hochgurtel, M. *et al*. Target-induced formation of neuraminidase inhibitors from *in vitro* virtual combinatorial libraries. *Proc. Natl Acad. Sci. USA* **99**, 3382–3387 (2002).
43. Congreve, M. S. *et al*. Detection of ligands from a dynamic combinatorial library by X-ray crystallography. *Angew. Chem. Int. Ed. Engl.* **42**, 4479–4482 (2003).
44. Bourne, Y. *et al*. Freeze-frame inhibitor captures acetylcholinesterase in a unique conformation. *Proc. Natl Acad. Sci. USA* **101**, 1449–1454 (2004).
45. Lewis, W. G. *et al*. Click chemistry *in situ*: acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks. *Angew. Chem. Int. Ed. Engl.* **41**, 1053–1057 (2002).
46. Hajduk, P. J. *et al*. Identification of novel inhibitors of urokinase via NMR-based screening. *J. Med. Chem.* **43**, 3862–3866 (2000).
47. Nienaber, V. L. *et al*. Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nature Biotechnol.* **18**, 1105–1108 (2000).
An early example of X-ray crystallography-driven screening applied to the discovery of urokinase inhibitors by fragment optimization.
48. Lesuisse, D. *et al*. SAR and X-ray. A new approach combining fragment-based screening and rational drug design: application to the discovery of nanomolar inhibitors of Src SH2. *J. Med. Chem.* **45**, 2379–2387 (2002).
49. Lange, G. *et al*. Requirements for specific binding of low affinity inhibitor fragments to the SH2 domain of (pp60)Src are identical to those for high affinity binding of full length inhibitors. *J. Med. Chem.* **46**, 5184–5195 (2003).
50. Hajduk, P. J. *et al*. Design of adenosine kinase inhibitors from the NMR-based screening of fragments. *J. Med. Chem.* **43**, 4781–4786 (2000).
51. Hopkins, A. L, Groom, C. R & Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **9**, 430–431 (2004).
This short article describes a simple but useful way to quantify 'ligand efficiency' (the contribution of each non-hydrogen atom in a lead to the overall binding affinity).
52. Erlanson, D. A., McDowell, R. S. & O'Brien, T. Fragment-based drug discovery. *J. Med. Chem.* **47**, 3463–3482 (2004).
53. Teague, S. J. Implications of protein flexibility for drug discovery. *Nature Rev. Drug Discov.* **2**, 527–541 (2003).

Acknowledgements

H. Jhoti for scientific input, R. Taylor for figure 1 and H. Sore for proof reading.

Competing interests statement

The authors declare competing financial interests: see Web version for details.

Online links

DATABASES

The following terms in this article are linked online to:
EntrezGene:
Caspase-3 | c-SRC | cyclin-dependent kinase-2 | p38 MAP kinase | protein tyrosine phosphatase-1B | tryptase | thymidylate kinase | urokinase

Access to this interactive links box is free online.