Crystallization is a very powerful technique for investigating matter at atomic resolution. Structural information from small molecules to biological macromolecules may be achieved at a high level of detail, allowing the determination of the short- and long-range chemical interactions in molecules and the study of the function of the proteins in living organisms. Relevant experimental and computational efforts are necessary to achieve this result, which requires multidisciplinary skills. All the steps involved in the route for crystal structure determination are discussed and the state-of-the-art and expected future improvements described.

Experimental steps

Crystallization

First, a crystal of the protein to be studied has to be formed. This is a necessary step to obtain an interference effect from a large assembly of molecules, all having the same configuration and put on the same orientation and position in the crystal. Unfortunately, this is very hard to achieve for protein structures, for a number of reasons. The protein sample present in the initial solution needs to be monodisperse, highly pure and concentrated at the desired value (typically 20 mg/ml for soluble proteins). With membrane proteins, additional difficulties arise from the fact that their hydrophobic surface hampers the formation of inter-molecular contacts needed for crystal growing. Therefore, detergent additives have to be added in the protein solution to favor the formation of protein crystals [1]. One could easily realize that the crystallization step represents the bottleneck for the whole protein structure determination pathway. Specific crystallization procedures have, thus, been developed to grow protein crystals, the most common of which rely on vapor diffusion. Specific crystallization procedures have, thus, been developed to grow protein crystals, the most common of which rely on vapor diffusion. They are designed so that the protein solution follows the path in the phase space illustrated in Figure 1: a transition from a soluble to a supersaturated state of the protein leads to the formation of the first crystal nuclei, occurring in the nucleation zone. Consequently, the protein concentration decreases and the crystal dimension increases during the transition toward the growth zone, located between the soluble and the nucleation regions. Crystallization takes place in sealed wells, arranged in plates similar to that represented in Figure 2. In each well, a volume of precipitant solution, called reservoir, and a droplet consisting of protein and precipitant solutions are present. Since the precipitant concentration is higher in the reservoir than in the drop, over time the water diffuses from the drop to the reservoir. Hence, the protein concentration in the drop increases up to double its initial value, and a specific path similar to that illustrated in Figure 1 can be followed, depending on the particular conditions in which the experiment has been carried out. As a matter of fact, the described process is long and difficult, because many efforts have to be made to establish

Protein crystallography and fragment-based drug design

Crystallography is a major tool for structure-driven drug design, as it allows knowledge of the 3D structure of protein targets and protein–ligand complexes. However, the route for crystal structure determination involves many steps, some of which may hamper its high-throughput use. Recent efforts have produced significant advances in experimental and computational tools and protocols. They include automatic crystallization tools, faster data collection devices, more efficient phasing methods and improved ligand-fitting procedures. The timescales of drug-discovery processes have been also reduced by using a fragment-based screening approach. Herein, the achievements in protein crystallography over the last 5 years are reviewed, and advantages and disadvantages of the fragment-based approaches to drug discovery that make use of x-ray crystallography as a primary screening method are examined. In particular, in some detail, five recent case studies pertaining to the development of new hits or leads in relevant therapeutic areas, such as cancer, immune response, inflammation, metabolic syndrome and neurology are described.
the proper conditions that allow the evolution of the drop solution from the soluble region to the labile region through the supersaturated one. Typical experimental parameters that are varied are: pH; initial precipitant concentration; initial protein concentration; temperature; and amount of solution put in the reservoir and in the drop. The type of precipitant, buffer and additives used for the drop and reservoir solutions are additional variables to be taken into account. Since a systematic exploration of all of them is out of question, they are usually sampled by a ‘sparse matrix’ approach. In designing crystallization experiments, more plates are considered, and the conditions explored in each plate form a 2D matrix, in which two variables are varied row- and column-wise and the most probable condition is put in the center (Figure 2). Many improvements have been achieved in recent years, in order to speed up the crystallization process, to enhance its efficiency and to reduce the amount of protein needed. The setting up of crystallization trials, their monitoring and classification have been automated, by using miniaturized crystallization plates and robot systems [2,3], and advanced crystallization kits are commercially available, to allow a better exploration of the parameter space. In addition, alternative crystallization techniques have been developed, which aim at improving the control on the process and its reproducibility. For example, a porous membrane to regulate the flux of vapor between drop and reservoir (membrane crystallization) [4], 3D nano-templates whose surface mesoporosity and surface chemistry is tuned on the protein of interest [5], or an intense electric field [6,7] were found effective in enhancing crystal growth.

- Data collection
Diffraction data are taken by irradiating the crystal sample with an intense and focused x-ray beam and collecting the diffracted rays by area detectors. x-ray sources for protein crystallography may be home diffractometers equipped with high-power rotating anode generators, or, better, synchrotron facilities. The availability of a high flux of incident x-ray photons is a crucial requirement, needed to compensate the poor diffracting properties of protein crystals. In fact, the latter are mainly composed of solvent and solutes (the solvent content may range from 20 to 80% of their volume), so that the protein units they contain have a residual flexibility, which hamper the Bragg scattering due to disorder. A fundamental parameter characterizing the quality of the diffraction data is the experimental resolution, which quantifies the amount of data collected, that is, the number of reflection intensities that could be recorded. An example of the effect of different experimental resolutions on the electron density map that can be obtained from them is provided in Figure 3. The experimental resolution can be improved by using x-ray detectors which have large sensitive areas, low noise, extended dynamical range and reduced radiation damage upon continuous data collection. Last-generation detectors, made by arrays of millions of silicon pixel channels combined by a CMOS readout chip, hold all these qualities [8,9] and, in addition, have very fast readout [10], so that the collection time has been dramatically decreased (one data set can be collected in minutes). Another beneficial way of collecting better diffraction data is to use cryogenic conditions, which has the effect to reduce the radiation damage on the crystal and the thermal motion of the protein atoms, so that higher data resolution can be reached. In view of reducing the data collection time, synchrotron facilities have been equipped by automatic sample loaders and even by protocols to perform remote experiments. In fact, the crystals can be now be sent to the synchrotron facility and the data collection can be monitored online through specific software tools [11].
Computational steps

Initial phasing
The outcome of the diffraction experiment consists of a set of measured intensities, one for each reflection ‘h’, representing a given direction of the scattered x-ray beam (scalar quantities and modulus of vector quantities are put in italics). The intensity, $I_h$, is proportional to the modulus of the structure factor, $F_h$, which represents the response of the crystal system to the x-ray perturbation. It is a complex quantity, which can be written as:

$$F_h = F_0 \exp(i\varphi_h) = \sum_{i=1}^{N} f_i \exp(i\mathbf{h} \cdot \mathbf{r}_i)$$

where $N$ is the number of atoms in the crystal cell (including symmetry equivalents), $f_i$ is the scattering factor of the $i$-th atom and $\mathbf{r}_i$ is its positional vector in fractional coordinates. However, to rebuild the image of the crystallized compound and then determine its structure, the phase values, $\varphi_h$, of the structure factors are also necessary. Thus the phase problem consists in determining the phase of structure factors from their moduli. It can be solved by different computational methods, called ‘phasing methods’ [12]. Their efficiency mainly depends on the quality of the diffraction data, that is, on the number of the intensities $I_h$ measured (data resolution) and on their accuracy. If the measured intensities are the only information available, the phasing methods are called ab initio and the classical ones are direct or Patterson methods. The first works in the reciprocal space (the space of the reflections ‘h’) and makes use of the quantities $F_h$ and $\varphi_h$; the latter operates in the direct space (the space of the coordinates ‘r’) and makes use of the electron density maps, $\rho(r)$. It should be noted that the direct and reciprocal spaces are dual spaces and that their related quantities are connected by the real part of the Fourier transform according to the following relations:

$$F_h = T[\rho(r)]$$

$$\rho(r) = T^{-1}[F_h]$$

where $T$ and $T^{-1}$ denotes the Fourier transform and antitransform, respectively. If, instead, additional information is available, such as a structural model of a homologous protein, the molecular replacement (MR) method can be applied to solve the phase problem [13]. It consists of finding the orientation and position of the model within the crystal cell of the target structure, and then in refining the correctly located model against the experimental data. Almost 90% of the 77,878 structures deposited in the Protein Data Bank (correct as of November 2012), which is the unique worldwide repository for macromolecules structures, were solved by x-ray diffraction and 60% of them were solved by MR. The number of solved macromolecular structures is increasing constantly. As the structural database increases in size, the MR method is used more frequently to derive the initial phase estimates for a new structure.

Phase refinement & model building
Phasing methods provide an initial estimate of the phases, which are used in Equation 2A to

Figure 2. Set up for crystallization experiments by vapor diffusion technique. (A) Single crystallization wells are arranged in plates, (B) where the crystallization conditions are explored row- and column-wise. (C) Each well is filled by precipitant and protein solutions.

Figure 3. Effect of finite data resolution. (A) Diffraction spots as measured by area detectors are grouped according to their experimental resolution, and (B) associated to the resolution of the electron density map that can be obtained at the end of the phasing process.
obtain an approximated electron density map. The phase values are then refined by specific iterative procedures, to obtain better estimates, hence, a more detailed electron density map, which can be interpreted and converted in a structural model. They are called electron density modification (EDM) procedures, and consist of modifying the electron density map by selecting its higher (positive) part and/or by powering it \([14,15]\). In the case of proteins, more advanced map modifications are performed, such as the definition of the solvent region \([16]\) and the adaptation of the pixel intensities to empirical pixel intensity distributions derived from a large set of known structures \([17]\). As a result, the Fourier inversion of the modified map \(\rho^*\) is expected to produce better phase values than those used for calculating the original map. It should be noted that in EDM the phase values are updated in each cycle, while the moduli are always set to their experimental values. This restrain is responsible for the convergence of the procedure towards more reliable phase estimates. At the end of the EDM cycles, the final electron density is automatically interpreted: atomic positions are associated to the higher peaks, since the peak height is proportional to the atomic Z-number, and their species are related to the peak height. A plausible structural model is then built by connecting the atoms found, consistently with stereochemical rules and with the known protein sequence. The structural model is finally refined by least squares procedure against diffraction data.

**Ligand fitting**

Once the protein model is reconstructed, eventual bound ligands may be identified by using the difference Fourier synthesis with the known protein structural model as reference. It can be written as:

\[
\rho_{\text{diff}}(t) = T^{-1} [(F_t - F_C) e^{\text{eff}}]
\]

**Equation 3** where \(F_t\) is the structure factor calculated from the protein structural model, and \(\rho_{\text{diff}}(t)\) represents the difference electron density map. Contrary to the standard electron density map \(\rho(t)\) of **Equation 2B**, which is expected to be definite positive, \(\rho_{\text{diff}}(t)\) holds valuable information either it is positive or it is negative. Positive \(\rho_{\text{diff}}(t)\) values can be found in regions where the actual structural model lacks some atom, negative ones indicate errors in the position or species assignment of the model atoms, or that the atom is not present at all in the target structure \([18]\).

An example is given in **Figure 4**, where \(\rho(t)\) is compared with \(\rho_{\text{diff}}(t)\) in a region with a ligand is expected to be present. When a large positive bump in \(\rho_{\text{diff}}(t)\) indicates the presence of a ligand, as in **Figure 4B**, the procedure of fitting the ligand in the electron density map can be initiated. The steps involved are the following:

- Define the ligand structure and its flexibility. Usually rotatable bonds are identified and their dihedral angles are left free to be varied;
- Global minimization algorithms, such as simulated annealing or genetic algorithm, are employed to fit the flexible ligand in the positive part of \(\rho_{\text{diff}}(t)\). Here, the values of the six parameters defining the position and orientation of the ligand, together with the values of its free torsion angles, are determined by using a proper cost function. It depends on the agreement between \(\rho_{\text{diff}}(t)\) and the electron density map calculated from the ligand. Additional terms are added, which implements stereochemical restraints for ligand–ligand and ligand–protein atoms;
- The protein–ligand structure undergoes least-squares refinement, by maximizing the agreement between observed and calculated structure factors.

Automatic procedures that implement the above steps have been developed \([19,20]\). Some of them are also able to fit in sequence a series of possible ligands and rank them, in analogy with protocols for docking a ligand into the active site \([21]\). Automatic ligand fitting is problematic in case of low occupancy of the ligand, for example, when the ligand has bound only a small fraction of all protein units present in the crystal. In these difficult cases, the discrimination of the cost function could be improved by using two recent theoretical

---

**Figure 4.** Electron density and difference-electron density maps give different details of the region occupied by the ligand. (A) \(\rho(t)\) and (B) \(\rho_{\text{diff}}(t)\) in a region where a ligand is present.
advancements. One is a tool to improve the quality of the different electron density map, based on the use of the EDM technique, has been recently proposed [22], and the other is a theory to calculate the error associated to the values of an electron density map has been recently developed [23].

**Fragment-based approach**

Protein crystallography is used in the framework of fragment-based drug discovery (FBDD), having the advantage of defining ligand-binding sites in protein–ligand complexes with more certainty [24]. For this reason all steps in the protein crystal structure solution have been automated.

The use of protein crystallography for studying the binding modes of molecular fragments, as a means of identifying new hits, started in 2000 [25]. It was pioneered at Abbott [101] and, since then, several companies, including Astex Therapeutics [102] and SGX Pharmaceutical [103], have used this technique to develop clinical candidates. A fragment-based screening (FBS) approach has been thus conceived, whose main differences with the high-throughput screening (HTS) approach are illustrated in Figure 5. The library of compounds to be screened in HTS is huge (approximately 10^8), since it is formed by the synthetic products obtained starting from a number (approximately 10^4) of small fragments. Moreover, the chance to obtain a successful binding of a library compound with the protein depends on the conformational flexibility of the given compound (it can explore a limited conformational space to adapt to the protein binding sites).

Different fragments selection strategies to build a good fragment library for FBS have been reported in literature [26]. In general, when designing a suitable fragment library, the main criteria relate to:

- Range of physicochemical properties of the fragments to be included;
- Aqueous solubility, which is required for screening;
- Molecular diversity;
- Chemical tractability of the fragments for follow-up;
- Chemical functionalities to be avoided (i.e., reactive compounds and known toxicophores);
- Drug-likeness of the fragments;
- Sampling of privileged scaffolds found in natural products and current drugs.

A fragment should have a MW ranging between 150 and 300 Da, and a high water solubility needed during screening, taking into account that fragments should be screened at high concentrations (0.2–1.0 mM) in aqueous buffer. Fragments to be included in a library are selected by applying a number of filters to large in-house or commercial compound collections and databases. Astex Therapeutics [27], for example, proposed the ‘role of three’ (MW ≤ 300 Da, cLogP ≤ 3, hydrogen bond donors ≤ 3, hydrogen bond acceptors ≤ 3, number of rotatable bonds ≤ 3), which is a variant of the well-known Lipinski’s ‘rule of five’ guidelines for oral drugs. To prioritize fragments with a good chance of binding to a protein target, further selection approaches, including in silico approaches (e.g., virtual screening), can be employed.

FBS proved to allow the generation of drug-like leads against difficult targets, reducing the time required to reach patent applications [28]. In addition, FBS libraries can adequately cover large chemical spaces with a reduced number of entries compared with HTS libraries, increasing the probability of finding hits [29]. Despite the low-binding affinity, crystallography allows specifically bound fragments to be visualized and their interactions defined.

**Cocktail crystallography**

Fragments can be studied for binding to their target proteins by using two approaches: co-crystallization and soaking. These two
Co-crystallization, the fragments are included in the crystallization solution together with the protein, so that crystals of the protein–ligand complex can be formed. In soaking, a preformed protein crystal is soaked in a solution containing the fragments. These fragments can rapidly diffuse into the protein crystal and interact with protein units as if they were in solution thanks to the extensive solvent channels present in protein crystals. Soaking is preferred to co-crystallization for high-throughput applications, since only the crystallization protocol for the apo-protein needs to be optimized. Co-crystallization is used as an alternative for the cases in which the crystal packing occludes the binding site (crystals of protein–ligand complexes can grow under slightly different conditions from those required for the protein alone).

Following the common protocol, apo-protein crystals are immersed into cocktails of molecules (up to 100), for a screening of a complete library of fragments (generally up to 1000; hence the term cocktail crystallography). Diffraction experiments on the soaked crystals allow identifying the fragments that bind to the target protein and studying their binding modes. These selected fragment hits can be optimized to leads and/or clinical candidates, allowing fragments to evolve from weak binders to ligands with nanomolar affinity. Fragment binding to the target can be improved by suitable substitution at one or more sites of a fragment hit with additional functional groups. For success of this approach it is required that the fragment hit acts as an ‘anchor’ that does not change its binding mode during its evolution to a potent lead. When more than one fragment has been detected that binds to different pockets of a target protein, they may be connected either directly or through suitable linkers, which allow the fragments to retain their original binding modes. Literature reports that fragment linking may be achieved by ‘click chemistry’ if there are conditions for mild chemistry (i.e., room temperature and aqueous environment).

A further successful strategy is the so-called ‘fragment tethering’, which relies on the formation of a disulfide bond between a reactive fragment and a cysteine residue in the target protein. Replacement of the disulfide bond with a methylene linker and further optimization with a more rigid linker has been observed to improve reversible ligand binding.

**Advantages & disadvantages of crystallography**

Applied to FBS, protein crystallography has contributed significantly to drug discovery and design, offering a number of advantages to the medicinal chemist:

- Binding modes are detected with high level of detail, which makes developing leads from fragments easier. The experimental resolution depends on the actual protein crystal and experimental apparatus used, but it usually ranges from 1 to 3 Å;
- False positives are not an issue, as fragments with no specific binding do not contribute to the electron density map (only structural features that are stable within the data acquisition time can be recovered by x-ray crystallography);
- Weaker binding fragments can be identified, which might be missed by other approaches [30]. In such cases, however, the ligand has low crystallographic occupancy, hence, additional efforts are required to reconstruct it in the electron density map;
- Large target proteins can be studied, since their structure solution is relatively easy [31]. Current phasing methods are not limited by the size of the protein, rather by the quality of diffraction data acquired (experimental resolution, mosaicity, redundancy and so forth);
- Mixture can be screened by soaking, so that the binding modes of several different ligands can be checked at the same time.
Besides the above advantages, there are some disadvantages to be considered, such as:

- The throughput is not very high, primarily because of the requirement for a crystallized protein;
- The binding affinities cannot be quantitatively assessed (no inhibition data can be determined);
- False negatives are possible. In fact, fragments may be discarded because their entrance in the protein crystal is hampered by the crystal packing. This could be avoided by using the co-crystallization technique, at the cost of a lower-throughput process;
- It requires specialized instrumentations (diffractometers), which are unlikely to be available to individual laboratories or large facilities (synchrotrons), which can be far from the laboratory.

**Role of crystallography in FBDD**

FBDD has significantly developed in the last 10 years [32], with more than 430 papers published. The number of publications proving the importance of protein crystallography in the advancement of FBDD, both in industry and academy, has increased accordingly (trend represented in Figure 7).

Vemurafenib (marketed as Zelboraf™), a B-Raf enzyme inhibitor developed by Plexxikon and Hoffman-La Roche for the treatment of late-stage melanoma [33], was the first US FDA-approved drug [34] discovered by FBDD (see below, case study 5).

As retrieved by SciFinder®, in substantial agreement with other tools of literature retrieving (PubMed, Scopus and ISI web of knowledge™), the papers reporting protein x-ray crystallography as a primary or secondary screening technique in FBDD studies are more than 130 (approximately 30% of the publications in the FBDD field).

A recent review article by Astex group [26] examined in some detail representative complexes between fragment-sized ligands and protein targets deposited in the PDB between 2004 and 2008, whereas successes and pitfalls in FBS by protein crystallography have been most recently reviewed by Chilingaryan and colleagues [35]. Herein, the focus will be on the developments in protein crystallography, as applied in FBDD studies, over the past 5 years. In Table 1, a number of representative examples of ligand-bound proteins are listed in descending chronological order, along with information about key biological functions of the target proteins and their importance as drug targets, and the PDB accession codes [36–67]. In some reported examples, drug leads (or even candidates) have been developed from fragments through FBDD, whereas in others cases fragments have been identified later by lead deconstruction. Below the authors will focus on five case studies in which x-ray crystallography has actually contributed, as screening technique, to improve FBDD. The fragment/ligand 2D structures of the remaining cases listed in Table 1 are represented in the Supplementary Data.

**Case 1: ultra-high-resolution structure of the carbohydrate-protein recognition domain in human Gal3C**

Galectins regulate cellular trafficking of glycoproteins, signaling and cell adhesion, thereby playing roles in cell growth, differentiation, cycle regulation and apoptosis [68]. A number of studies have demonstrated the potential of high-affinity and selective galectin inhibitors in modulating inflammation processes, immune response and cancer growth [69–72]. In a very recent study [44], the ultra-high-resolution crystal structures of the carbohydrate-binding site in Gal3C, both in the ligand-free state (at 100 and 298 K) and in complex with lactose or glycerol, have been reported. The x-ray crystallographic study was usefully complemented with NMR spectroscopy, molecular dynamics...
## Table 1. Examples of ligand-bound proteins whose crystal structures have been recently reported in fragment-based drug-discovery studies†.

<table>
<thead>
<tr>
<th>Target protein</th>
<th>Bound compound/s, PDB code/s</th>
<th>Importance as drug target</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDHA</td>
<td>1, 4AJ2; 2, 4AJ1; 3, 4AJ4; 4, 4AJH; 4, 4AJK; 5, 4AJL; 6, 4AJE; 7, 4AJI; 8 + 4, 4AJJ; 9, 4AJN; 10, 4AL4; 11, 4AJP; 11, 4AJO</td>
<td>LDHA, a tetrameric enzyme existing in five isoforms, is a key glycolytic enzyme that catalyzes the conversion of pyruvate to lactate, utilizing NADH as co-factor. Given that it is often upregulated in tumors and its high expression is linked to poor prognosis, LDHA has been thought of as a therapeutic target for cancer therapy</td>
<td>[36]</td>
</tr>
<tr>
<td>Pim-1 kinase</td>
<td>12, 3VBW; 13, 3VBV; 14, 3V8X; 15, 3VBY; 16, 3VB; 17, 3VCA; 18, 3VBQ</td>
<td>Belonging to the Pim family of Ser/Thr kinases, it plays a functional role in cell survival, demonstrating potential as target in oncology and immunology regulation</td>
<td>[37]</td>
</tr>
<tr>
<td>Bromodomain/ AcK pocket</td>
<td>19, 4A9P; 20, 4A9O; 21, 4A9N; 22, 4A9M</td>
<td>Bromodomains are epigenetic reader modules that regulate the structure of chromatin, and thereby gene transcription, through their recognition of acetyl-lysine modified histone tails. Inhibitors of this protein–protein interaction have the potential to modulate multiple diseases as demonstrated by the anti-inflammatory and anti-proliferative effects</td>
<td>[38,39]</td>
</tr>
<tr>
<td>ROCK</td>
<td>23, 3V85</td>
<td>ROCKs belong to the AGC (PKA/PKG/PKC) family of Ser/Thr kinases. They are downstream effector proteins of the small GTPase Rho, which is one major regulator of the cytoskeleton. ROCK1 and 2 enzymes are involved in several pathological conditions, such as cardiovascular diseases, glaucoma, inflammatory disorders and cancer</td>
<td>[40]</td>
</tr>
<tr>
<td>Hsp90</td>
<td>24, 2XDK; 25, 2XDL; 26, 2XDU; 27 + 26, 2XDS; 28, 2XDX; 29, 2XHR; 30, 2XHT; 31, 2XHX; 32, 2XAB</td>
<td>Interference with Hsp90 and its associated clients can allow the simultaneous targeting of the hallmarks of cancer (e.g., insensitivity to apoptotic signaling, angiogenesis, replication, invasion and metastasis). Preclinical and clinical data support the potential utility of Hsp90 inhibitors in anticancer therapy</td>
<td>[41]</td>
</tr>
<tr>
<td>Hsp90</td>
<td>33 (co-crystal), 2W1Z; 33 (soak), 2W1B; 34, 2W1T; 35, 2W14; 36, 2W15; 37, 2W16; 38, 2W17</td>
<td>Interference with Hsp90 and its associated clients can allow the simultaneous targeting of the hallmarks of cancer (e.g., insensitivity to apoptotic signaling, angiogenesis, replication, invasion and metastasis). Preclinical and clinical data support the potential utility of Hsp90 inhibitors in anticancer therapy</td>
<td>[42]</td>
</tr>
<tr>
<td>Hsp90</td>
<td>39, 4EEH; 40, 4EFT; 41, 4EFU</td>
<td>Interference with Hsp90 and its associated clients can allow the simultaneous targeting of the hallmarks of cancer (e.g., insensitivity to apoptotic signaling, angiogenesis, replication, invasion and metastasis). Preclinical and clinical data support the potential utility of Hsp90 inhibitors in anticancer therapy</td>
<td>[43]</td>
</tr>
<tr>
<td>Gal3C (Case study 1)‡</td>
<td>lactose, 3ZSJ; glycerol, 3ZSK; apo at 100K, 3ZSL; apo at 298 K, 3ZSM</td>
<td>Galectins are small soluble proteins involved in regulation of cellular trafficking of glycoproteins, signaling and cell adhesion. They are potential targets in inflammation, immunity and cancer, due to their roles in cell growth, differentiation, cycle regulation and apoptosis</td>
<td>[44]</td>
</tr>
<tr>
<td>Cytochrome bc1 complex</td>
<td>42, 3TGU</td>
<td>As membrane protein, the cytochrome bc1 complex is a central component of the cellular respiratory chain of the photosynthetic apparatus in photosynthetic bacteria a promising target for antiparasitic agents, antibiotics and fungicides</td>
<td>[45]</td>
</tr>
<tr>
<td>PDE4A</td>
<td>43, 3VTX</td>
<td>PDE4, which consists of four subfamilies (A–D), is a cAMP-specific phosphodiesterase for which selective inhibitors have been studied as anti-inflammatory agents for the treatment of asthma, chronic obstructive pulmonary disease and rheumatoid arthritis</td>
<td>[46]</td>
</tr>
<tr>
<td>PDK1</td>
<td>44, 3PWW; 45, 3QC4</td>
<td>PDK1 phosphorylates highly conserved Ser or Thr residues in the T-loop of at least 23 related downstream AGC protein kinases. PDK1 signaling pathway is critical for cancer cell growth, survival and tumor angiogenesis. Its inhibitors could provide a therapeutic cancer treatment</td>
<td>[47]</td>
</tr>
</tbody>
</table>

†The structures of the fragment ligands are reported in Figs. 9, 11, 13 & 15A (case studies 2, 3, 4 and 5, respectively) and in the Supplementary Material.
‡The case studies discussed in the text.

uPA: Urokinase-type plasminogen activator.
### Table 1. Examples of ligand-bound proteins whose crystal structures have been recently reported in fragment-based drug-discovery studies (cont.)*

<table>
<thead>
<tr>
<th>Target protein</th>
<th>Bound compound/s, PDB code/s</th>
<th>Importance as drug target</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDK1</td>
<td>46, 3SC1</td>
<td>PDK1 phosphorylates highly conserved Ser or Thr residues in the T-loop of at least 23 related downstream AGC protein kinases. PDK1 signaling pathway is critical for cancer cell growth, survival and tumor angiogenesis. Its inhibitors could provide a therapeutic cancer treatment</td>
<td>[48]</td>
</tr>
<tr>
<td>p38 (MAPK)</td>
<td>47, 3P5K; 48, 3P7B; 49, 3P7A; 50, 3P7B; 51, 3P7C; 52, 3P79</td>
<td>Plays a key role in the regulation of the cytokines TNFα and IL-1β. As such, p38 MAPK has potential as a target for the development of inhibitors of inflammatory diseases.</td>
<td>[49]</td>
</tr>
<tr>
<td>Endothiapepsin</td>
<td>53 (ritonavir), 3PRS; 54 (saquinavir), 3PWW</td>
<td>Member of the aspartic protease class of enzymes that are widely distributed among fungi, plants, vertebrates and viruses. HIV protease is essential for maturation of the virus particle and its inhibitors have a proven therapeutic utility in the treatment of AIDS. Aspartic proteases also play major roles in hypertension, amyloid disease and malaria, and have been implicated in tumorigenesis.</td>
<td>[50]</td>
</tr>
<tr>
<td>HK (Case study 2)</td>
<td>55 + 56, 3NBV; 57, 3NBW; 58, NCA; 59, 3NC2; 60, 3NC9</td>
<td>KHK catalyzes, with ATP and K+, the conversion of α-fructose to fructose-1-phosphate, initiating the intracellular catabolism of dietary carbohydrates, showing potential as a drug target for metabolic syndromes (e.g., obesity, Type 2 diabetes)</td>
<td>[51]</td>
</tr>
<tr>
<td>HK (Case study 2)</td>
<td>61, 3RO4</td>
<td>KHK catalyzes, with ATP and K+, the conversion of α-fructose to fructose-1-phosphate, initiating the intracellular catabolism of dietary carbohydrates, showing potential as a drug target for metabolic syndromes (e.g., obesity, Type 2 diabetes)</td>
<td>[52]</td>
</tr>
<tr>
<td>AChBP</td>
<td>62, 2Y54; 63, 2Y56; 64, 2Y57; 65, 2Y58</td>
<td>Water-soluble pentameric protein homolog of the ligand binding domain of Cys-loop receptors, such as nicotinic acetylcholine receptors, GABA, 5-HT3 and glycine receptors</td>
<td>[53]</td>
</tr>
<tr>
<td>Bovine pancreatic</td>
<td>66, 3RXA; 67, 3RXB; 68, 3RXC; 69, 3RXD; 70, 3RXE; 71, 3RXF; 72, 3RXG; 73, 3RXH; 74, 3RXI; 75, 3RXJ; 76, 3RXL; 77, 3RXK; 78, 3RXM; 79, 3RXO; 80, 3RXP; 66 + 70, 3RXQ; 66 + 68, 3RXR; 69 + 74, 3RXS; 69 + 68, 3RXT; 70 + 74, 3RXU; 70 + 68, 3RXV; 69 + 66, 3ATI; 66 + 74, 3ATK; 74 + 69, 3ATL; 74 + 68, 3ATM</td>
<td>Serine protease that cleaves peptide chains mainly at the carboxyl side of Lys and Arg residues. The role of trypsin in pancreatic diseases, including cystic fibrosis and chronic pancreatitis, is currently investigated. It has been used to model the decomposition of articular cartilage in osteoarthritis</td>
<td>[54]</td>
</tr>
<tr>
<td>Bovine pancreatic</td>
<td>81, 3NKK; 82, 3NKB</td>
<td>Serine protease that cleaves peptide chains mainly at the carboxyl side of Lys and Arg residues. The role of trypsin in pancreatic diseases, including cystic fibrosis and chronic pancreatitis, is currently investigated. It has been used to model the decomposition of articular cartilage in osteoarthritis</td>
<td>[55]</td>
</tr>
<tr>
<td>LTA4H</td>
<td>83, 3FTS; 84, 3FTU; 85, 3FTX; 86, 3FTV; 87, 3FTW; 88, 3FTY; 89, 3FTO; 90, 3FUB; 91, 3FU5; 92, 3FUI; 93, 3FUD; 94, 3FUI; 95, 3FUE; 96, 3FUF; 97, 3FUJ; 98, 3FUJ; 99, 3FUK; 100, 3FUM; 101, 3FUN</td>
<td>LTA4H catalyzes the final step in the synthesis of leukotriene B4. While earlier studies of LTA4H inhibitors focused on allergy, dermatitis and arthritis, the inhibitor DG-051 recently entered Phase II clinical trials in cardiovascular diseases</td>
<td>[56]</td>
</tr>
<tr>
<td>LTA4H</td>
<td>102, 3FHE; 103, 3FH8; 104, 3FTZ; 105, 3FUL; 106, 3FH5; 107, 3FH7</td>
<td>LTA4H catalyzes the final step in the synthesis of leukotriene B4. While earlier studies of LTA4H inhibitors focused on allergy, dermatitis and arthritis, the inhibitor DG-051 recently entered Phase II clinical trials in cardiovascular diseases</td>
<td>[57]</td>
</tr>
<tr>
<td>hPNMT (Case study 4)</td>
<td>108, 3KP; 109, 3KPU; 110, 3KPV; 111, 3KPW; 112, 3KPY; 113, 3KQM; 114, 3KQO; 115, 3KQP; 116, 3KQQ; 117, 3KQ; 118, 3KQ; 119, 3KQV; 120, 3KQW; 121, 3KQY; 122, 3KRO; 123, 3KR1; 124, 3KR2</td>
<td>Catalyzes the final step in adenylate biosynthesis and is involved in a range of biological processes and diseases, including the central control of blood pressure, pituitary hormone secretion, ethanol intoxication, Parkinson’s and Alzheimer’s diseases</td>
<td>[58]</td>
</tr>
</tbody>
</table>

*The structures of the fragment ligands are reported in Figures 9, 11, 13 & 15A (case studies 2, 3, 4 and 5, respectively) and in the Supplementary Material.

†The case studies discussed in the text.

uPA: Urokinase-type plasminogen activator.
<table>
<thead>
<tr>
<th>Target protein</th>
<th>Bound compound/s, PDB code/s</th>
<th>Importance as drug target</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BACE1</td>
<td>125, 3HW1; 126, 3HVG</td>
<td>Aspartyl-protease enzyme, whose inhibition may have utility in the treatment of Alzheimer's disease, catalyzes the rate-limiting step in the production of the β-amyloid peptide and amyloid plaque formation in Alzheimer's disease</td>
<td>[59]</td>
</tr>
<tr>
<td>CMX-T β-lactamase</td>
<td>127, 3G2Y; 128, 3G2Z; 129, 3G30 (1.80); 130, 3G31; 131, 3G32; 132, 3G34; 133, 3G35</td>
<td>Extended-spectrum class A β-lactamase, that confers bacterial resistance to third-generation cephalosporins</td>
<td>[60]</td>
</tr>
<tr>
<td>JAK-2</td>
<td>134, 3E62; 135, 3E63; 136, 3E64</td>
<td>As a cytoplasmic receptor-associated protein Tyr kinase of the JAK family, it is implicated in the response to receptors from the single-chain and IL-3 cytokine families, as well as the IFN-γ receptor. Selective inhibition of a constitutively active JAK-2 may have potential toward myeloproliferative disorders</td>
<td>[61]</td>
</tr>
<tr>
<td>uPA</td>
<td>137, 2VIN; 138, 2VIO; 139, 2VIP; 140, 2VIQ; 141, 2VIN; 142, 2VIW</td>
<td>Trypsin-like serine protease uPA catalyzes the conversion of plasminogen to plasmin. It is implicated in a variety of disease states associated with abnormal tissue destruction and cell infiltration, including aortic aneurism, multiple sclerosis and metastasis</td>
<td>[62]</td>
</tr>
<tr>
<td>B-Raf kinase</td>
<td>143, 3D4Q</td>
<td>The high prevalence of B-Raf mutations in human cancer and their role in the Ras/Raf/MEK/ERK signaling pathway in tumorigenesis and progression supports its potential as drug target in cancer chemotherapy</td>
<td>[63]</td>
</tr>
<tr>
<td>B-Raf kinase</td>
<td>144, 3C4E; 145, 3C4F; 146, 3C4C; 147, 3C4D (4FK3)</td>
<td>The high prevalence of B-Raf mutations in human cancer and their role in the Ras/Raf/MEK/ERK signaling pathway in tumorigenesis and progression supports its potential as drug target in cancer chemotherapy</td>
<td>[64]</td>
</tr>
<tr>
<td>B-Raf kinase</td>
<td>148, 3OG7</td>
<td>The high prevalence of B-Raf mutations in human cancer and their role in the Ras/Raf/MEK/ERK signaling pathway in tumorigenesis and progression supports its potential as drug target in cancer chemotherapy</td>
<td>[65]</td>
</tr>
<tr>
<td>DPP-4</td>
<td>149, 3CCB; 150, 3CCC</td>
<td>DPP-4 is a serine protease that rapidly cleaves and inactivates the incretin GLP-1 in the blood, which ultimately enhances glucose-stimulated insulin release from beta cells in the pancreas. Its inhibitors (e.g., sitagliptin) are effective agents in the treatment of Type 2 diabetes</td>
<td>[66]</td>
</tr>
<tr>
<td>PGDS</td>
<td>151, 2VD0; 152, 2VCW; 153, 2VVC; 154, 2VCX; 155, 2VCQ; 156, 2VD1</td>
<td>PGDS, which exists in two isoforms, catalyzes the isomerization of PGH₂ (i.e., the common precursor of various PGs) to PGD₂, whose release results in diverse biological responses (e.g., sleep promotion, inhibition of platelet aggregation, attraction of inflammatory cells and bronchoconstriction). Its inhibitors (e.g., tranilast) are used in the treatment of allergic disorders</td>
<td>[67]</td>
</tr>
</tbody>
</table>

†The structures of the fragment ligands are reported in Figures 9, 11, 13 & 15A (case studies 2, 3, 4 and 5, respectively) and in the Supplementary Material.

The case studies discussed in the text.

uPA: Urokinase-type plasminogen activator.
simulations and isothermal titration calorimetry (ITC). The apo- and ligand-bound crystal structures revealed that water and carbohydrate oxygen atoms are in similar positions in all three states. Lactose binds Gal3C (resolution of 0.86 Å) in a conformation similar to that observed in previously reported crystal structures for \( \text{N}-\text{acetyllactosamine} \) and lactose as well, at lower resolution [73–75]. Conserved water molecules making important bridging hydrogen bonds and Gal3C key residues, mainly His158, Arg160, Asn174 and Glu184, which significantly contribute to the binding of lactose to Gal3C, are highlighted in Figure 8. Deuterium NMR relaxation dispersion experiments and molecular dynamics simulations provided complementary information on the role of water molecules and hydrogen-bonding networks at the Gal3C carbohydrate-binding site.

Overall, this study highlights the role of protein cryo-crystallography in screening carbohydrate lead fragments and investigating carbohydrate–protein interactions, which underlie numerous biological functions and are involved in a number of diseases, including inflammation, virus infection and cancer [76–78]. As regards cancer, glucose transporters are overexpressed in malignant cells [79], such as colon, breast and lung carcinomas.

Very recently, in a study aimed at exploring fragment-growing effects in new anticoagulant agents, some of us synthesized and tested for enzyme inhibition activity novel \( \beta\text{-d-glucosyl} \) conjugates of highly potent inhibitors of blood coagulation factor Xa and thrombin [80]. The fragment deconstruction of the sugar-bearing compounds led to quantify the binding affinity increase, which resulted in a small, but significant, \( \Delta G \) gain (2.9-3.8 kJ·mol\(^{-1}\)) for the C3-alkyl-linked glucose fragment.

**Case 2: discovery & optimization of novel indazole-based inhibitors of KHK**

Researchers at Johnson & Johnson described a FBDD approach to the identification and optimization of novel indazole-containing KHK inhibitors [51,52]. KHK, which catalyzes, with ATP and K\(^+\), the conversion of \( \text{d-fructose} \) to fructose 1-phosphate, is implicated in regulation of dietary sugar metabolism. The product of the KHK-catalyzed reaction, that is fructose 1-phosphate, enters the glycolytic pathway to provide a significant supply of carbon for the biosynthesis of fatty acids and very LDLs. Therefore, inhibition of KHK activity is thought to provide therapeutic benefit in non-insulin-dependent (Type 2) diabetes, obesity or hypertension.

The reported FBDD protocol [51] consists of three iterations of design, synthesis and x-ray crystallographic screening, which lead to submicromolar KHK inhibitors with promising drug-like properties starting from low-molecular-weight fragments. Using electron density of fragments within the protein binding site as defined by x-ray crystallography, target specific leads were generated without the use of affinity data from enzyme assays. The primary library is composed of approximately 900 fragments with no target class biases and target privileged functional groups, nor pharmacophores. Filters applied in composing the primary library were: number of non-hydrogen atoms \( \geq 6 \) and \( \leq 15 \), number of hydrogen bond acceptors \( \leq 3 \), number of hydrogen bond donors \( \leq 3 \), number of rings \( \leq 2 \) and no unspecified chiral centers. Moieties containing reactive, toxicophore or nondrug-like groups were removed and the fragments retained in the primary library only if they appeared as substructures in compounds collected in the Comprehensive Medicinal Chemistry database, a collection of over 8000 known pharmaceutical compounds [104]. A secondary library was designed based on hits found from the primary library, which led to identify six scaffolds: that is pyrazoles; arylimidazoles; benzotriazoles; pyridines; sulfones; and triazolones.Hits were discovered for compounds elaborated on the first four scaffolds. While a
fragment-growing approach was used to prepare the second library, fragment fusion was used in building up the tertiary library. The second and tertiary libraries consist each of 300–500 compounds. 2D structures of the small molecules crystallized in complex with KHK during these studies are represented in Figure 9. Adenylyl imidodiphosphate (55), as a stable ATP analogue and D-fructose (56) bind the catalytic site of KHK C crystal (PDB code: 3NBV; 2.3 Å resolution). Hit indazole compound 60, as a structural evolution of the primary library pirazole 59, was reported to bind the ATP-binding site with submicromolar binding affinity and good PK properties.

The molecular optimization process, assisted by both x-ray crystallography (Figure 10) and enzyme activity data, led to compounds with high inhibition potency and good PK profiles.

The indazole-containing compound 61 (PDB code: 3RO4; binding mode in Figure 10) was found to inhibit KHK with IC\textsubscript{50} of 0.59 µM, and clean selectivity profile toward numerous biological targets (G-protein-coupled receptors, ion channels, transporters and enzymes). In a rat PK study it demonstrated good oral bioavailability (98.8%), good oral exposure in plasma (AUC = 5679 ng h/µl), moderate clearance (28.9 ml/min/kg) and volume of distribution at steady state (7.81 l/kg).

**Case 3: ranking of fragments hit compounds with trypsin inhibitory activity through crystallographic competition experiments**

In this approach bovine pancreatic trypsin has been used as target protein to highlight a relationship between the in-crystal binding affinities of hit compounds found in the Fo-Fc electron density maps and their inhibitory activities in solution [54].

After hit-compound identification by x-ray-based screening, crystallographic competition experiments were carried out to determine the relative order of the in-crystal binding affinities.

The focused primary fragment library contained 62 simple molecular scaffolds bearing amino (primary, secondary), hydroxyl and carboxyl groups and seven compounds known to bind trypsin, forming hydrogen bonds with Asp189 at the bottom surface of the S1 pocket. A total of 15 compounds (66–80; Figure 11) were identified as trypsin binders by x-ray crystallographic screening. They were then categorized into five classes, based on the amino-group type and scaffold feature, and five representative hit compounds were selected from each scaffold class on the basis of structural simplicity, for crystallographic competition experiments: cycloheptanamine (66, scaffold I), pyridine-2-amine (68, scaffold II), benzamidine (70, scaffold III), [3-(aminomethyl)
A total of 10 mixed solutions of all possible combinations of two hit compounds, and soaking experiments were carried out to investigate the relative order of the binding affinities in the protein crystals (in-crystal affinity ranking). The results of the competition experiments confirmed that clear $F_o - F_c$ maps of the singlet fragment compound were observed in all 10 cases. The pairwise examination of the crystallographic results led to establish an in-crystal affinity ranking of the five representative hits that was in reasonable agreement with the rank of inhibitory activities as assessed in solution by a fluorescence-based bioassay. Moreover, the correlation analysis revealed that x-ray-based screening could detect a weak hit compound (68) with inhibitory activity below the limit of detection of the high-concentration enzyme assay. Figure 12 shows the $2F_o - F_c$ electron density map contoured at 2σ for the better binder fragment (i.e., benzamidine 70; PDB: 3RXE), as well as the overlay of the fragments 66, 69, 70 and 74, soaked with bovine pancreatic trypsin.

This study demonstrates the successful assessment of in-crystal affinity ranking through crystallographic competition experiments, thereby expanding the potential of the protein x-ray crystallography as a primary screening method in FBDD. However, as the authors stated, the in-crystal affinity ranking cannot be considered of general applicability. Further studies have to be undertaken to establish if and how the in-crystal affinity ranking by competition studies are influenced by conformational changes.
Case 4: FBS of human PNMT inhibitors

The enzyme PNMT, which catalyzes the methylation of noradrenaline to adrenaline (the final step of the adrenaline biosynthesis), could have a potential role in a number of CNS pathologies, including the central control of blood pressure, pituitary hormone secretion, ethanol intoxication, Parkinson's and Alzheimer's diseases. Structures of the recombinant human PNMT (hPNMT) in complex with inhibitors and substrates (e.g., PDB code 1HNN) demonstrated that hPNMT has an enclosed active site, suggesting that enzyme conformational changes may be needed for allowing ligand binding \[\text{[81]}\]. The apo-enzyme does not readily crystallize in the absence of AdoHcy (S-adenosyl-L-homocysteine; 108; \text{Figure 13}), which binds to the co-factor-binding site, and, moreover, when crystals are grown in the absence of a substrate or inhibitor at the noradrenaline-binding site, this site is occupied by a phosphate molecule \[\text{[82]}\].

In a recent study, Martin et al. successfully applied FBS approaches to identifying PNMT inhibitors by using x-ray crystallography as the primary screen \[\text{[58]}\]. A total of 12 fragment hits (109–120), which include benzimidazoles, purines and quinolines, were identified from a commercial library of 384 drug-like fragments, and their binding affinities were determined by ICT. Nine of the hits had measurable binding affinities.
Five further benzimidazole compounds (121–124 and the derivative bearing NH₂, Cl, OH at the positions 2, 5 and 7, respectively, for which high-resolution diffraction data were not collected) were synthesized and their binding to PNMT determined by ITC (Kᵣ = 1.8–20 μM; LE = 0.58–0.72), showing for the first time how this class of inhibitors interact with the noradrenaline-binding site. In Figure 14, the x-ray crystal structure of benzimidazole compound 123 (Kᵣ = 1.8 μM; LE = 0.72) soaked with hPNMT is represented.

In a further study by Nair et al. [83], a cocktail containing resorcinol, imidazole and 6-chloroindolin-2-one (123) has been investigated for their binding to hPNMT, showing how theoretical approaches, including MD simulations and free energy calculations, can be used to detect cooperative binding and to validate potential hits. Co-ordinates of PNMT complexed with resorcinol, imidazole and AdoHcy (108) have been deposited (PDB code 4DM3). To test whether resorcinol and imidazole can bind stably to hPNMT, 5-ns simulations were performed within the binding pocket. Interestingly, as a major outcome, resorcinol adopted a stable binding mode, whereas imidazole remained in the pocket but adopted a number of different binding modes during the simulation. In contrast, in presence of the resorcinol fragment, the binding mode of imidazole was stable and perpendicular to the plane of resorcinol. In summary, this study demonstrated that computational approaches can be helpful in identifying stable tertiary complexes allowing the x-ray data from cocktails to be exploited.

**Case 5: fragment-based discovery of vemurafenib, a 7-azaindole inhibitor of oncogenic B-Raf kinase activity for the treatment of late-stage melanoma**

The finding that V600E is the most common mutation in the B-Raf oncogene in the majority of melanomas has stimulated the search for potent and selective inhibitors of B-RafV600E kinase activity [64]. Researchers at Plexikon screened a library of 20,000 compounds (MW range 150–350 Da) against a number of diverse kinases, and 238 compounds were found to inhibit three kinases (Pim-1, p38 and CSK) at 200 μM concentration. More than 100 crystal complex structures were solved, which revealed the binding of 7-azaindole to the ATP-binding site of Pim-1. A follow-up screen using x-ray crystallography allowed a group of monosubstituted 7-azaindoles to be identified, which includes compounds 144 and 145 displaying micromolar inhibition potency against and consistent binding modes to kinase domains of Pim-1 and FGFR1, respectively. Subsequent fragment optimization led to the discovery of PLX4720 (146), a derivative that inhibits B-RafV600E with an IC₅₀ value of 13 nM, and proved to be tenfold selective against wild-type B-Raf and more than 100-fold selective against a panel of 70 diverse other kinases. In B-RafV600E-dependent tumor xenograft models, orally administered PLX4720 demonstrated significant anti-melanoma activity.

Using a crystallography-guided approach, PLX4032 (148) was discovered which displayed similar potency for B-RafV600E (IC₅₀ = 31 nM), and selectivity against many other kinases, including wild-type B-Raf (100 nM) [65]. The 3D structure of compound 148 binding to B-RafV600E is represented in Figure 15.

Toxicological studies confirmed a wide margin of safety, whereas the Phase I clinical data demonstrated a high 81% response rate in the metastatic melanoma patients receiving an oral dose of 960 mg, twice daily. PLX4042, whose nonproprietary name ‘vemurafenib’ comes from V600E mutated B-Raf inhibition, has received approval for the treatment of late-stage melanoma in the USA, Canada and the EU [84,85].

**Future perspective**

Steps forward to overcome the limitations of protein crystallography for FBS have been made,
which include automation of crystal structure solution and ligand-fitting procedures, faster data collection, optimization of protocols for cocktail preparation and many others. Further advances could be achieved by using alternative methods for pre-screening the fragment library, so that only promising hits are analyzed by crystallography. An example of this approach has been given in the case studies examined herein.

**Box 1. Recent achievements of fragment-based crystallography.**

**Crystallization**
- Automatic crystallization tools [2,3]
- New crystallization methods [4–7]

**Data taking**
- Automatic sample exchanger and remote experiments [11]
- Higher intensity and focused x-ray beams
- Advanced detectors [8–10]

**Phasing**
- Automatic crystal structure solution [90–92]
- More efficient molecular replacement methods [93,94]

**Ligand fitting**
- Automatic software procedures [19,20]
- Difference electron density modification [22]
- Use of the map variance to improve the cost function [23]

**Fragment screening**
- Advanced cocktails
- Pre-screening by alternative methods [44,86–89]

The most used alternative methods include surface plasmon resonance [86], ITC [87], tethering [88] mass spectrometry, high concentration screening and computational methods [89]. The main achievements can be summarized in **Box 1**.

Possible alternative strategies include the application the fragment-based approach for multitarget drug design. In fact, one could be submit a set proteins crystallized in their apo form to the same ensemble of cocktails, so that the same fragment library is screened by several target proteins. After crystal structure solution, the diverse x-ray structures are compared, searching for conserved relative position of some fragments. These represent the starting points for rationally designing multitarget ligands by step-wise addition of linkers and/or functional groups [95].

### Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.Future-Science.com/doi/full/10.4155/FMC.13.84

### Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.
Recent technical and computational advances have increased the throughput level of the fragment-based crystallography. Four recent case studies highlight the increasing potential of the fragment-based crystallography as a primary screening method in drug design. Requires a number of experimental and computational steps.

A number of recent advancements in the crystallization methods and in their implementation have substantially speeded-up the process and enhanced its efficiency. Data taking at home diffractometers or at synchrotron sources is now a routine task, which can be also executed remotely.

Fragment-based approach

The key difference between high-throughput and fragment-based screenings is that in the former the fragments are combined before the crystallographic study, during the synthesis step, whereas, in the latter, they are combined after it. This leads to a drastic reduction of the library size and to an increased efficiency of the drug-design process.

Crystallographic fragment-based screening is implemented by using cocktails, that is, solutions containing a number of compounds when the protein crystals are soaked. The compounds that are found to bind the protein are used to develop hits.

Fragment-based crystallography

Crystallization of the protein sample is the most difficult and low-throughput step, since its output depends on a large number of variables. Alternative techniques can be used to pre-screen the fragment.

Protein crystallography is used in the framework of drug design to screen libraries of compounds. The compounds that bind to the protein can be studied with high structural details.

Experimental steps

Crystallization of the protein sample is the most difficult and low-throughput step, since its output depends on a large number of variables. A number of recent advancements in the crystallization methods and in their implementation have substantially speeded-up the process and enhanced its efficiency.

Data taking at home diffractometers or at synchrotron sources is now a routine task, which can be also executed remotely.

Computational steps

Diffracted x-rays cannot be focused as with light rays. The image of the protein structure can only be obtained by applying phasing procedures.

Ligands are reconstructed on the protein structure by means of automatic procedures that fit them in the electron density map.

Protein crystallography is used in the framework of drug design to screen libraries of compounds. The compounds that bind to the protein can be studied with high structural details.

Ligands are reconstructed on the protein structure by means of automatic procedures that fit them in the electron density map.

Crystallography offers unique advantages over alternative screening techniques. Data taking at home diffractometers or at synchrotron sources is now a routine task, which can be also executed remotely.

Crystallographic fragment-based screening is implemented by using cocktails, that is, solutions containing a number of compounds when the protein crystals are soaked. The compounds that are found to bind the protein are used to develop hits.

Fragment-based crystallography

The key difference between high-throughput and fragment-based screenings is that in the former the fragments are combined before the crystallographic study, during the synthesis step, whereas, in the latter, they are combined after it. This leads to a drastic reduction of the library size and to an increased efficiency of the drug-design process.

Crystallographic fragment-based screening is implemented by using cocktails, that is, solutions containing a number of compounds when the protein crystals are soaked. The compounds that are found to bind the protein are used to develop hits.

Fragment-based approach

Four recent case studies highlight the increasing potential of the fragment-based crystallography as a primary screening method in drug discovery and design.

References

Papers of special note have been highlighted as:

- of interest
- of considerable interest


7 Koizumi H, Uda S, Fujiiwara K et al. Control of effect on the nucleation rate for hen egg white lysozyme crystals under application of an external ac electric field Langmuir. 27 (13), 8333–8338 (2011).


Fundamental text to understand theoretical and practical aspects of crystallography.


• Theoretical bases of a new tool to improve the quality of the difference electron density map.


• Review highlighting trends and developments in fragment-based drug discovery, including representative protein–fragment complexes deposited in the Protein Data Bank between 2004 and 2008.


• Review highlighting and discussing successes and pitfalls in recent fragment-based screening by protein crystallography.


Successful application of fragment-based screening approaches to identifying inhibtors, by using x-ray crystallography as the primary screen.


Websites
102 AstexTherapeutics. Pharmaceutical company. www.astex-therapeutics.com
103 SGX Pharmaceutical. Pharmaceutical company www.lilly.com