

VIRTUAL SCREENING

finding needles in a haystack on a shoestring

In recent years, virtual screening (VS) has become an accepted approach to hit-finding alongside biochemical high-throughput screening (HTS). Large pharmaceutical companies are typically in the fortunate position of possessing sizeable historical compound collections or combinatorial libraries for use in VS and HTS campaigns. But these are not a pre-requisite for successful hit-finding. Here we present a highly efficient and cost-effective paradigm for hit-finding by VS that has been applied with success in a small company with, by comparison, relatively limited resources.

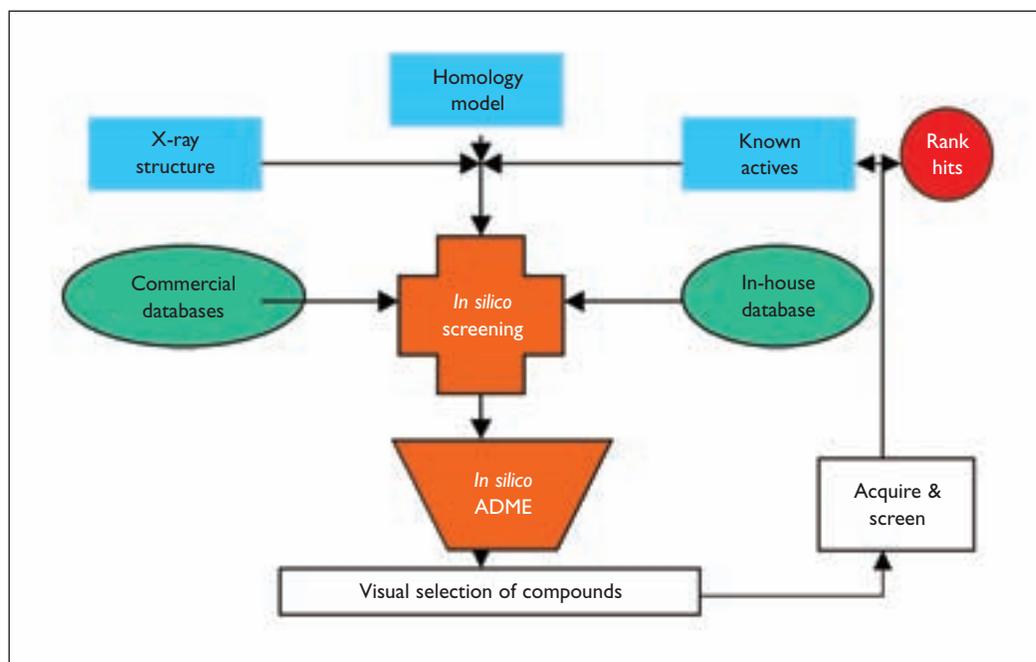
The use of computational methods to search databases of chemical structures for compounds similar to a known active molecule or to dock putative ligands into protein active sites is hardly new – both approaches have their roots in the 1980s, if not earlier^{1,2}. However, in recent years, renewed research effort has been poured into these *in silico* approaches as the promise of ‘virtual screening’ has begun to be realised. There is now a steady stream of publications that testifies to the efficiency and effectiveness of VS in the discovery of novel compounds active against a particular biological target³⁻¹⁰. Additionally, there is growing evidence that VS should be seen as complementary to, rather than in competition with, high-throughput biochemical screening^{11,12}.

Virtual screening techniques can be divided into two classes:

- Ligand-based – in which databases of chemical structures are interrogated to find compounds that are similar to known actives (similarity searching) or possess a pharmacophore or substructure in common with a known active (pharmacophore and substructure searching). Similarity and substructure searching may be carried out with reference to either the 2-D or the 3-D structure of a compound.
- Structure-based – in which rapid docking algorithms are used to place candidate compounds within the active site of the biochemical target of interest and then score them according to their steric and electrostatic complementarity to the site.

**By Dr David E. Clark,
Neil V. Harris,
Alan G. Roach and
Dr Anthony D.
Baxter**

Figure 1
Overview of virtual screening process



It is fair to say that there has been more interest in structure-based, rather than ligand-based, virtual screening approaches in recent years but this should not obscure the considerable power of the latter, which we shall illustrate later in this article.

Rapid Reaction Teams – a hit-finding paradigm

Early in its existence, Argenta Discovery developed a paradigm for hit-finding based on virtual screening. The rationale for this was several-fold. First, the company was a start-up, so possessed no historical compound collection, though a modest collection of 25,000 commercial compounds was assembled early on to provide some screening capability. Second, relatively little medicinal chemistry resource was available for patent-busting approaches or synthesis of proprietary screening libraries. Third, the company needed a ‘fail-fast, fail-cheap’ approach since it could not afford to expend endless resources on a project unlikely to yield success in a short time-frame. So, the concept of Rapid Reaction Teams (RRT) was born.

The key features of the RRT approach are as follows:

- The aim is to complete hit-finding in six months. This will typically involve two rounds of virtual screening, the second following up the hits from the first to generate early structure-activity relationships. However, if very promising series are identified in the first round of screening, the

approach is flexible enough to allow an early transition to hit-to-lead chemistry.

- The approach is driven by virtual screening coupled with biochemical screening. Little medicinal chemistry resource is required.

- A starting point is required – either a known active compound, or several of them, or an x-ray structure of the biochemical target. Where a target structure and known ligands are available, both ligand-based and structure-based VS will be carried out.

Underpinning the virtual screening strategy is Argenta’s database of commercially available screening compounds and a collection of virtual screening tools. The database is collated from more than 40 suppliers worldwide and is updated three times a year to ensure currency. In its most recent incarnation, the database, after filtering using a battery of drug-likeness parameters, comprises over a million compounds. For structure-based virtual screening, Argenta possesses the FlexX¹³ automated ligand docking programme together with the CScore¹³ consensus scoring package. A variety of tools is available for ligand-based virtual screening including 2-D similarity searching based on atom-pair descriptors¹⁴, atom environment descriptors¹⁵ and Daylight and Unity fingerprints^{13,16}; 2-D and 3-D substructure searching using Unity and Daylight SMARTS^{13,16} and 3-D similarity searching using FlexS¹³.

The overall virtual screening process is summarised in **Figure 1**. Starting from the available information, virtual screening is used to identify a set of compounds (typically a few thousand) for further evaluation. This assessment includes a battery of *in silico* ADME filters and the judgement of an experienced medicinal chemist. In this way, as much care as possible is taken to ensure that any actual hits are 'drug-like' and also attractive starting points for further chemical exploration. The selected compounds, usually fewer than 1,000, are then obtained from the appropriate vendors and screened in the biochemical assay. Hits are validated by obtaining solid samples and establishing proof of identity by NMR and purity by LC/MS.

The timelines and sequence of activities for a typical RRT project are shown in **Table 1**.

Typically, the total resource applied over the six-month period will not exceed three FTEs, ie the hit-finding exercise is completed in less than 1.5 person years. Clearly, the RRT approach is not only rapid, but also highly cost-effective, compared to a high-throughput screening campaign, which may involve the screening of hundreds of thousands of compounds, probably at the cost of hundreds of thousands of dollars.

The proof of the pudding

The RRT approach has become Argenta's main method of hit-finding for its proprietary therapeutics research and has also been used successfully for projects for clients. Some examples of applications of the RRT paradigm in Argenta's therapeutics programmes are summarised in **Table 2**.

Our first three attempts at the application of

this methodology met with mixed success. For the histone deacetylase (HDAC) inhibitor programme we used only structure-based virtual screening (using an x-ray structure of a homologous protein), and we physically screened just 30 compounds in order to identify multiple (modestly) potent hits. Optimisation of one these series (selected on the bases of chemical tractability, novelty etc) has led to the discovery of novel, potent HDAC inhibitors. Our second attempt at an enzyme target (Enzyme Inhibitor 1) was less successful. Both structure- and ligand-based approaches were used for this programme, and again a number of modestly potent inhibitors were identified, but all of these compounds proved to have non-competitive enzyme kinetics (our project objective was to discover competitive inhibitors). However, our third programme, and the first directed towards a GPCR target (MCH-1 receptor antagonists), was very successful. Again we used both structure- and ligand-based approaches. The structure-based approach used a model of the receptor based on the published structure of rhodopsin, and there were at least 11 known antagonists at this receptor that we could use for the ligand-based approach. In practice, the structure-based approach failed to identify any hits at all, whereas the ligand-based methods led to the discovery of multiple hits, some of which had IC_{50} values in the binding assay of less than 100nM (some of this work has been published^{17,18}). The failure of the structure-based approach for this GPCR target is perhaps not surprising. While there is some emerging evidence that homology models of GPCRs can be successfully used in structure-based virtual screening^{19,20}, it is still the

References

- 1 Kuntz, ID, Blaney, JM, Oatley, SJ, Langridge, R, Ferrin, TE. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 1982, 161:269-288.
- 2 Jakes, SE, Watts, N, Willett, P, Bawden, D, Fisher, JD. Pharmacophoric pattern matching in files of 3D chemical structures: evaluation of search performance. *J Mol Graphics* 1987, 5:41-48.
- 3 Lyne, PD, Kenny, PVW, Cosgrove, DA, Deng, C, Zabludoff, S, Wendoloski, JJ, Ashwell, S. Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge-based virtual screening. *J Med Chem* 2004, 47:1962-1968.
- 4 Wang, S, Meades, C, Wood, G, Osnowski, A, Anderson, S, Yuill, R, Thomas, M, Mezna, M, Jackson, W, Midgley, C, Griffiths, G, Fleming, I, Green, S, McNae, I, Wu, SY, McInnes, C, Zheleva, D, Walkinshaw, MD, Fischer, PM. 2-Anilino-4-(thiazol-5-yl)pyrimidine CDK inhibitors: synthesis, SAR analysis, X-ray crystallography, and biological activity. *J Med Chem* 2004, 47:1662-1675.
- 5 Singh, J, Chuaqui, CE, Boriack-Sjodin, PA, Lee, WC, Pontz, T, Corbley, MJ, Cheung, HK, Arduini, RM, Mead, JN, Newman, MN, Papadatos, JL, Bowes, S, Josiah, S, Ling, LE. Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGFbeta receptor kinase (TbetaRI). *Bioorg Med Chem Lett*. 2003, 13:4355-4359.
- 6 Peng, H, Huang, N, Qi, J, Xie, P, Xu, C, Wang, J, Yang, C. Identification of novel inhibitors of BCR-ABL tyrosine kinase via virtual screening. *Bioorg Med Chem Lett*. 2003, 13:3693-3699.
- 7 Vangrevelinghe, E, Zimmermann, K, Schoepfer, J, Portmann, R, Fabbro, D, Furet, P. Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *J Med Chem* 2003, 46:2656-2662.

DISCIPLINE	MONTH					
	1	2	3	4	5	6
CADD	Virtual screen round 1. Order compounds		Receive compounds	Virtual screen round 2. Order compounds	Receive compounds. Order solid samples	Receive solid samples
Biochemistry	Develop and validate primary screen (and functional assay) Screen round 1			Screen round 1	Screen round 2	Verify assay results on solid samples
Medicinal Chemistry	Synthesise standards. Aid selection of compounds		Plate out compounds		Plate out compounds	Verify structures of solid samples

Table 1: Timelines and sequence of activities for a typical RRT project

Continued on page 40

Continued from page 39

8 Schapira, M, Raaka, BM, Das, S, Fan, L, Totrov, M, Zhou, Z, Wilson, SR, Abagyan, R, Samuels, HH. Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. *Proc Natl Acad Sci U S A* 2003, 100:7354-7359.

9 Varady, J, Wu, X, Fang, X, Min, J, Hu, Z, Levant, B, Wang, S. Molecular modeling of the three-dimensional structure of dopamine 3 (D3) subtype receptor: Discovery of novel and potent D3 ligands through a hybrid pharmacophore- and structure-based database searching approach. *J Med Chem* 2003, 46:4377-4392.

10 Flohr, S, Kurz, M, Kostenis, E, Brkovich, A, Fournier, A, Klabunde, T. Identification of nonpeptidic urotensin II receptor antagonists by virtual screening based on a pharmacophore model derived from structure-activity relationships and nuclear magnetic resonance studies on urotensin II. *J Med Chem* 2002, 45:1799-1805.

11 Paiva, AM, Vanderwall, DE, Blanchard, JS, Kozarich, JW, Williamson, JM, Kelly, TM. Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of *Mycobacterium tuberculosis*. *Biochim Biophys Acta* 2001, 1545:67-77.

12 Doman, TN, McGovern, SL, Witherbee, BJ, Kasten, TP, Kurumbail, R, Stallings, WC, Connolly, DT, Shoichet, BK. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem* 2002, 45:2213-2221.

13 FlexX, CScore, Unity and FlexS are all available from Tripos, Inc. www.tripos.com

14 Carhart, RE, Smith, DH, Venkataraghavan, R. Atom pairs as molecular features in structure-activity studies: definition and applications. *J Chem Inf Comput Sci* 1985, 25:64-73.

Table 2: Applications of RRT paradigm at Argenta Discovery

PROJECT	STARTING POINTS AND APPROACH	NUMBER OF COMPOUNDS SCREENED	RESULTS	ELAPSED TIME FOR HIT-FINDING
Histone deacetylase inhibitor	X-ray structure of homologous protein with known inhibitor bound Structure-based virtual screening	30	Compounds with IC ₅₀ values in the range 1-20µM discovered; one series selected for optimisation	~6 months
Enzyme inhibitor I	X-ray structure of homologous protein with known inhibitor bound; two competitor inhibitors Structure- and ligand-based virtual screening	225	Six inhibitors discovered with IC ₅₀ values in the range 20-30µM; all inhibitors showed non-competitive kinetics	4 months
MCH-1R antagonist	Homology model of receptor plus structures of 11 competitor compounds Ligand-based virtual screening	795	Seven distinct structural series discovered, with compounds that gave IC ₅₀ values in the range 0.055-27µM	6 months
GPCR1 agonist	Single competitor compound Ligand-based virtual screening	220 (+643*)	At least six distinct structural series discovered, with compounds that gave K _i values in the range 0.032-2.5µM	4 months
GPCR2 antagonist	Three known antagonists Ligand-based virtual screening	643 (+220*)	Two distinct structural series discovered – one hit compound equipotent in binding assay to competitor molecule	4 months
GPCR3 (+GPCR4*) antagonist	Two known antagonists Ligand-based virtual screening	844	Two distinct structural series discovered – one hit compound equipotent to competitor molecule	3 months
* see text				

Continued on page 41

Continued from page 40

case that the retinal-bound x-ray structure of rhodopsin is not an ideal template for modelling GPCRs^{21,22}, particularly when the ligands are agonists. Additionally, the construction of an accurate, high-quality homology model is a time-consuming process and so would not fit well within the RRT approach described here. Consequently, we now prefer to leave the building of homology models until later in the project when they can be applied to the rationalisation of SAR within specific series.

In order to build on the success achieved with the MCH-1R programme, we decided to concentrate next exclusively on GPCR targets that appeared to us to have good starting points for the ligand-based methodologies. In these later projects we also tried to improve the efficiency of the whole process further by selecting multiple targets from the same gene family. The wisdom of this choice was borne out by the results we achieved. For example, the GPCR1 and GPCR2 targets described in Table 2 come from the same gene family, although we required agonists for GPCR1 and antagonists for GPCR2. For GPCR1 there was only one known small molecule ligand in the literature, and using this as a template yielded a screening set of 220 compounds, from which we discovered one hit series. However, by screening the set of compounds selected for the related GPCR2 target against GPCR1, we identified other hit series, despite the fact that the virtual screening set for GPCR2 was constructed based on ligands that were known to be antagonists at GPCR2, whereas the required function at GPCR1 was for agonism (and all of the compounds we found active in the GPCR1 binding assay proved to be agonists).

Similarly, for target GPCR3 there were only two known ligands in the literature, one of which was selective for GPCR3, the other of which was also active against the closely related receptor GPCR4. Ligand-based searches using known ligands for both GPCR3 and GPCR4 led to the discovery of compounds that were active against GPCR3. In fact, this screening exercise led to the discovery of novel compounds that were selective for both GPCR3 and GPCR4, as well as compounds that showed affinity at both receptors. As a further validation of our approach it should be noted that the most potent hits against GPCR1, GPCR2 and GPCR3 have *in vitro* profiles (affinity, potency, cell permeability, metabolic stability, aqueous solubility and effects on CYP450 isozymes) that largely match the profiles of the best compounds known in the literature (ie compounds that are in pre-clinical or even clinical development).

Summary

A rapid and cost-effective paradigm for hit-finding based upon virtual screening has been developed and validated at Argenta Discovery through application to several targets of therapeutic interest. Ligand-based virtual screening has proven to be particularly effective against GPCR targets finding, in a short space of time, potent and novel ligands that provide excellent starting points for subsequent medicinal chemistry programmes. **DDW**

David Clark is Director of Computer-aided Drug Design and Knowledge Management at Argenta Discovery. David obtained his PhD from the University of Sheffield. He was a founder member of Argenta in 2000 having previously worked for Proteus Molecular Design and Rhône-Poulenc Rorer/Aventis.

Neil Harris is a Director of Medicinal Chemistry at Argenta Discovery. Neil has more than 25 years' experience of medicinal chemistry, initially with the Rhône-Poulenc Group of Companies (1975-2000) and subsequently with Argenta Discovery, of which he was a founding member.

Alan Roach is Director of Therapeutics at Argenta Discovery. Alan started his research career in the pharmaceutical industry in 1974 by joining Synthélabo in Paris. Subsequent career moves brought him back to the UK with Reckitt & Coleman, Glaxo and to Rhône-Poulenc Rorer where he was recruited as Director of Vascular Biology in 1991. Between 1996 and 1997 he was in charge of Global Safety and General Pharmacology based in RPR's site at Vitry, France. From 1997 to the closure of the Aventis research site at Dagenham in the UK in July 2000, Alan worked in the Business Development & Licensing Group.

Anthony Baxter is Chief Executive Officer of Argenta Discovery. Before joining Argenta, Dr Baxter was Chief Scientific Officer with Oxford Asymmetry International (1995-2000) where he set up and managed the Discovery Services Division. Prior to OAI, he was Research Manager at Ciba's UK Central Research Laboratories (1990-1995) where he managed Ciba's 'blue-sky' research interests and before that he was Team Leader at Glaxo Group Research (1983-1990). Dr Baxter completed his PhD with Professor Stan Roberts on prostaglandin chemistry at Salford University.

15 Bender, A, Mussa, HY, Glen, RC, Reiling, S. Molecular similarity searching using atom environments, information-based feature selection, and a naive Bayesian classifier. *J Chem Inf Comput Sci* 2004, 44:170-178.

16 Available from Daylight Chemical Information Systems, Inc. www.daylight.com

17 Clark, DE, Higgs, C, Wren, SP, Dyke, HJ, Wong, M, Norman, D, Lockey, PM, Roach, AG. A virtual screening approach to finding novel and potent antagonists at the MCH-1 receptor. *J Med Chem*, accepted for publication.

18 Arienzo, R, Clark, DE, Cramp, S, Daly, S, Dyke, HJ, Lockey, P, Norman, D, Roach, AG, Stuttle, K, Tomlinson, M, Wong, M, Wren, SP. Structure-activity relationships of a novel series of melanin-concentrating hormone (MCH) receptor antagonists. *BioOrg Med Chem Lett*, accepted for publication.

19 Bissantz, C, Bernard, P, Hibert, M, Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? *Proteins* 2003, 50:5-25.

20 Evers, A, Klebe, G. Ligand-supported homology modeling of G-protein-coupled receptor sites: models sufficient for successful virtual screening. *Angew Chem Int Ed Engl* 2004, 43:248-251.

21 Furse, KE, Lybrand, TP. Three-dimensional models for beta-adrenergic receptor complexes with agonists and antagonists. *J Med Chem* 2003, 46:4450-4462.

22 Bissantz, C, Logean, A, Rognan, D. High-throughput modeling of human G-protein coupled receptors: amino acid sequence alignment, three-dimensional model building, and receptor library screening. *J Chem Inf Comput Sci* 2004, 44:in press.