ELSEVIER

A chemistry wiki to facilitate and enhance compound design in drug discovery

Graeme R. Robb, Darren McKerrecher, Nicholas J. Newcombe and Mike J. Waring

Cardiovascular and Gastrointestinal Innovative Medicines Unit, AstraZeneca R&D, Alderley Park, Alderley Edge SK10 4TG, UK

At AstraZeneca a focus on hypothesis-driven design and the formation of drug design teams has placed a greater emphasis on collaboration in the drug discovery process. We have created a novel software tool based on the principles of wikis and social networks to facilitate collaborative working, visual planning and incorporation of predictive science to improve design capability. Monitoring the design and make process via the tool enabled the identification of bottlenecks and delays. Solutions to these problems were implemented, reducing the time taken from the initial idea stage to the generation of the synthesised compound by more than 50%.

Introduction

In the recent past, drug discovery efforts, as exemplified by multiparallel synthesis (MPS) and HTS, have focused on making and testing ever increasing numbers of compounds. The unstated hypothesis underlying this strategy is: 'make and test enough compounds and the right one will be found'. However, drug-like chemical space is vast [1] and the probability of finding the 'right' compound, particularly when limited by chemistry techniques amenable to rapid and parallel compound synthesis, must be small. Recent analyses of the industry seem to confirm this [2–5]. The rate of successful drug discovery has not increased since the introduction of these techniques.

The alternative strategy is to focus on quality compound design, aided by all the experience and tools at our disposal. This more focused approach should not be limited by what might be considered 'easy' chemistry but should exploit advanced synthetic chemistry techniques to make the chosen compounds. The implication is that, with potentially longer synthesis routes per compound, fewer compounds will be made. This requires the compound designers to select clearly and prioritise the best compounds for synthesis. It also places a heavier burden on the design part of the design-make-test-analyse (DMTA) process, where it is imperative that expert designers (from different disciplines) collaborate and share knowledge effectively so that the greatest value can be derived from the data and the best compounds designed to reach a candidate drug in the fewest possible number of design iterations. At AstraZeneca drug designers have embraced these principles of hypothesis-driven design and have formed design teams of medicinal, computational, synthetic and physical chemists with pharmacokinetics experts on each project [6–8].

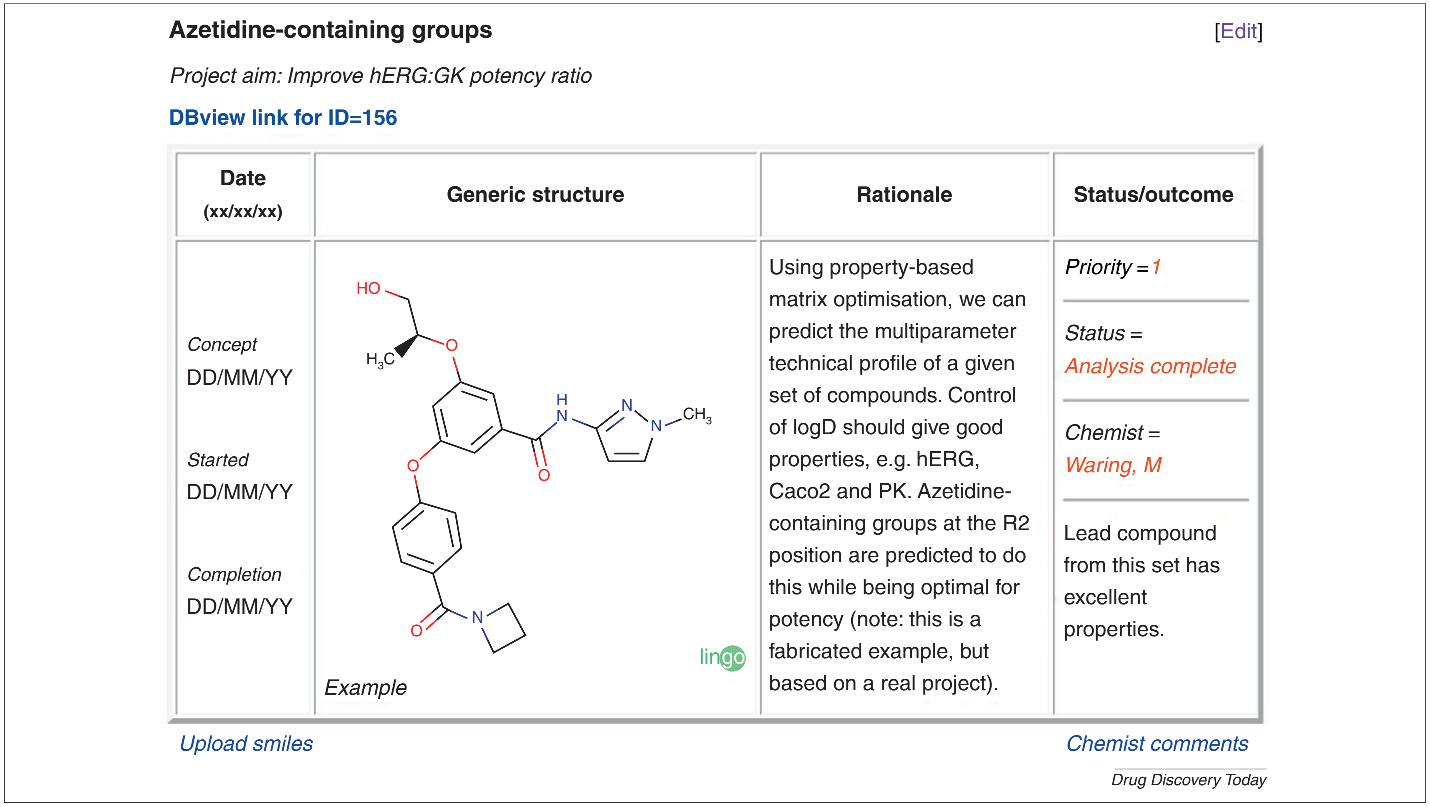
Working effectively as a team cannot be taken for granted, especially because design team members at AstraZeneca are not necessarily located together; projects often being split between research sites. External pressures on drug discovery mean that pharma companies are increasingly using services, including chemical synthesis, provided by contract research organisations [9,10]. Furthermore, even for the simplest case where a project exists at a single site and with in-house resource, the role of the designer (medicinal chemist) and the role of the maker (synthetic chemist) commonly mirror the outsourcing model and are kept separate, potentially creating a disconnect between these roles. These separations create additional challenges for design teams in collaborating efficiently.

Compound design tracking

It was quickly realised that in order for design teams to function effectively they required additional tools and infrastructure. Team members needed a means of recording and assessing each new design idea in a transparent fashion. They needed a platform for sharing files and data (linked to a design idea), so they could easily collaborate and they required a way to visualise the current design

Corresponding author:. Robb, G.R. (graeme.robb@astrazeneca.com), (graeme.r.robb@gmail.com)

1359-6446/06/$ - see front matter © 2012 Elsevier Ltd. All rights reserved. http://dx.doi.org/10.1016/j.drudis.2012.09.002



Azetidine-containing groups [Edit]

Project aim: Improve hERG:GK potency ratio

DBview link for ID=156

Date (xx/xx/xx)	Generic structure	Rationale	Status/outcome
Concept DD/MM/YY *Started* DD/MM/YY *Completion* DD/MM/YY	*Example*	Using property-based matrix optimisation, we can predict the multiparameter technical profile of a given set of compounds. Control of logD should give good properties, e.g. hERG, Caco2 and PK. Azetidine-containing groups at the R2 position are predicted to do this while being optimal for potency (note: this is a fabricated example, but based on a real project).	*Priority = 1* *Status = Analysis complete* *Chemist = Waring, M* Lead compound from this set has excellent properties.

Upload smiles *Chemist comments*

Drug Discovery Today

FIGURE 1

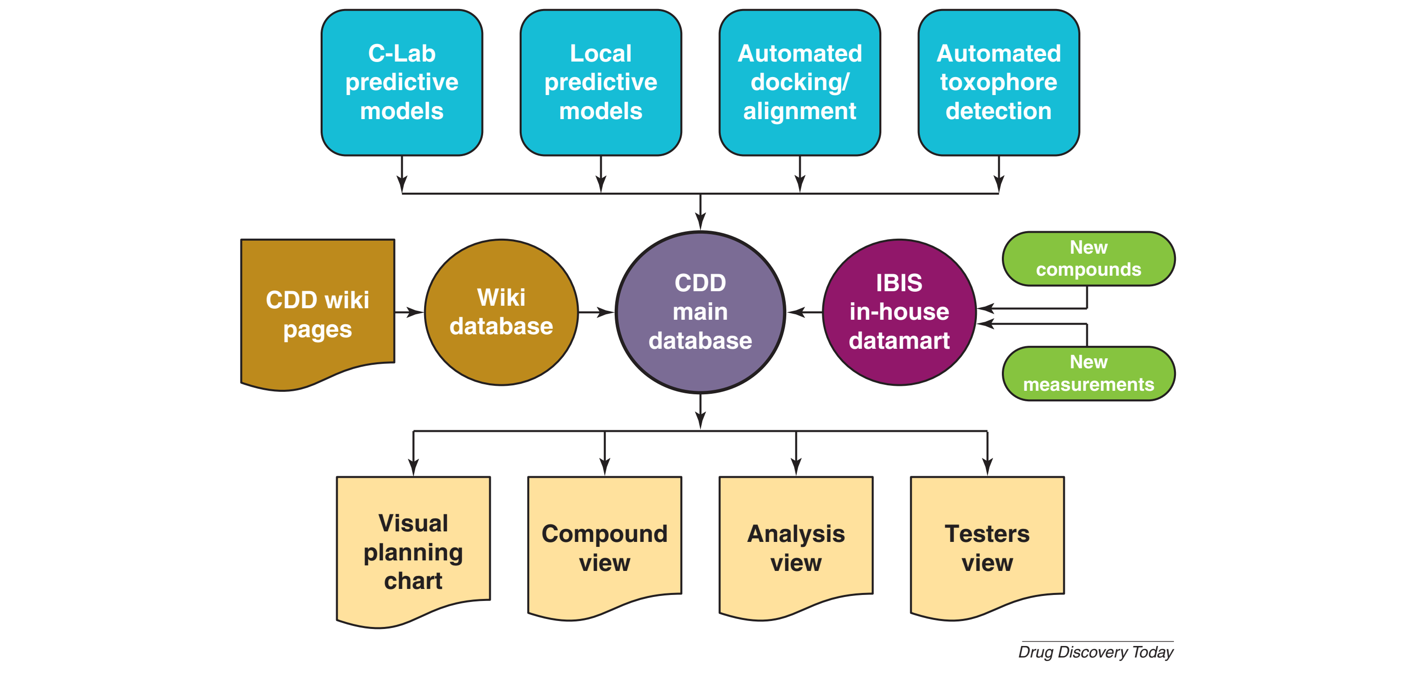
An example of a design idea as presented in the wiki environment. The various fields are completed within a template and hyperlinks are created to enable compounds or comments to be attached to this idea. Compound is AZD1092 [25].

ideas and their progress. Previous work in this field has highlighted the benefits of having a single tool to manage the DMTA workflow [11–13], of integrating data sources within a common interface [14,15] and of capturing medicinal chemistry knowledge for future design [16,17]. While learning from these examples, we opted for a novel and flexible solution to fit exactly the requirements of design teams at AstraZeneca.

Web2.0 technologies [18] and social media (as exemplified by Facebook, Wikipedia and Amazon) were deemed to offer the required features. These web-based solutions are transparent, always up-to-date, open for all users to view and contribute to, interoperative (i.e. they bring in data from multiple sources) and completely customisable (i.e. displaying the same data in various different ways to suit the task or purpose at hand). Moreover, by putting the tool at the heart of the process it was not seen as an overhead – something extra to be done and therefore neglected when time pressures were felt – but rather integral to the work, ensuring it remained useful and relevant. In terms of software development Web2.0 offers many advantages. There is no client-based software to write or support, updates take place on the server and there is no requirement to roll out updates to individual users. Additionally, production and pre-production versions can be maintained in parallel and rapid software revisions can be tested and then implemented with minimal impact on the user.

The prototype tool created was named the Compound Design Database [19]. It repurposed the mediawiki technology (as used for Wikipedia; http://www.mediawiki.org) to create design pages for each project. Simple templates were used to facilitate creation of new design ideas, where the designer could draw an example structure and enter details of their hypothesis. Chemoinformatics functionality was plugged into the wiki environment to provide true chemical structure understanding, rather than merely images. Molecular information was stored as SMILES. These were pasted directly or created by sketching molecules with the integrated java applet and the structure automatically displayed.

Initially projects simply used these pages as a focus for discussion and decision making. Keeping the open and inclusive nature of Web2.0, no user access restrictions were put in place – with all users free to contribute design to any project. The tool continued to evolve as projects used it, adding more fields to the template to capture the status of an idea and the priority as well as the ability to link files relevant to the design idea (Fig. 1). Utilising the architecture that underlies the wiki, the information was extracted from these pages and turned into a useable database. This greatly enhanced the potential of the tool and the ability to add lists of potential compounds for each design was added. Each compound could then have various metadata added, including date of creation, whether or not it already existed in the AstraZeneca compound collection (or similarity to its closest neighbour) and notes for synthesis such as reagent availability. Links were made to other databases and tools using molecular structure (encoded as a hash) as the unique key. This freed us from a reliance on any particular common language or technology between databases or tools. All relevant data were presented to the user or linked from a single hub, enabling informed decision-making (Fig. 2 for details of how the Compound Design Database is connected to other systems).



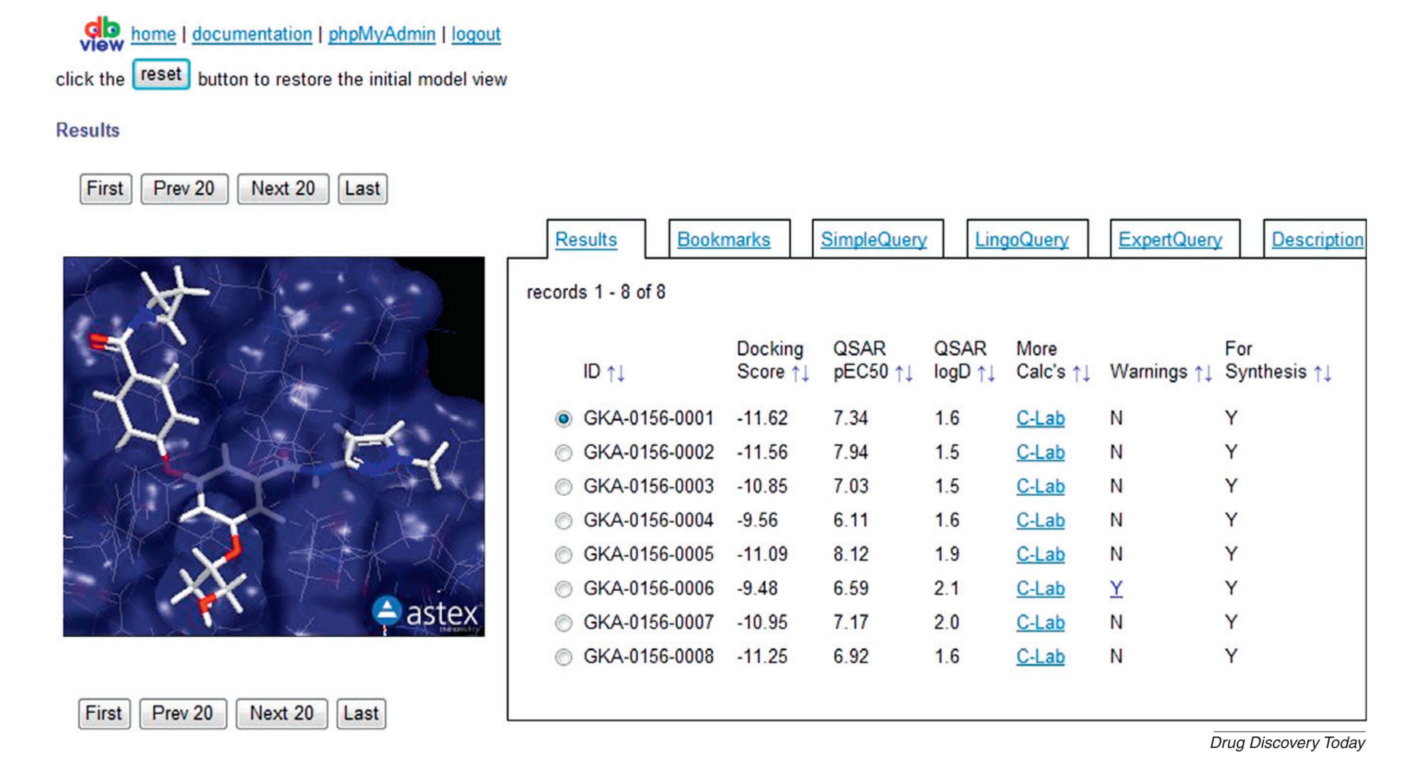
Drug Discovery Today

FIGURE 2

Information flow for the Compound Design Database (CDD) showing how data are collected from various sources and remixed in a variety of ways for output.

Enhancing compound design capability

It was quickly recognised that an opportunity existed to exploit this database containing new, and potentially soon-to-be-made, compounds further. Additional computational tools were brought to bear and predicted properties were computed for all compounds using robust global QSAR models, using AstraZeneca's C-Lab platform [20–22]. Other validated local QSAR models were used on a project-by-project basis to predict more properties of interest. Various 3D structural techniques were automatically run on compounds from the database, including protein docking and



click the reset button to restore the initial model view

Results

records 1 - 8 of 8

ID ↑↓	Docking Score ↑↓	QSAR pEC50 ↑↓	QSAR logD ↑↓	More Calc's ↑↓	Warnings ↑↓	For Synthesis ↑↓
● GKA-0156-0001	-11.62	7.34	1.6	C-Lab	N	Y
○ GKA-0156-0002	-11.56	7.94	1.5	C-Lab	N	Y
○ GKA-0156-0003	-10.85	7.03	1.5	C-Lab	N	Y
○ GKA-0156-0004	-9.56	6.11	1.6	C-Lab	N	Y
○ GKA-0156-0005	-11.09	8.12	1.9	C-Lab	N	Y
○ GKA-0156-0006	-9.48	6.59	2.1	C-Lab	Y	Y
○ GKA-0156-0007	-10.95	7.17	2.0	C-Lab	N	Y
○ GKA-0156-0008	-11.25	6.92	1.6	C-Lab	N	Y

Drug Discovery Today

FIGURE 3

An example of a custom view showing predicted properties and 3D visualisation of docking into the protein structure. Hyperlinks provide access to additional information.

pharmacophore similarity, with a new view to aid visualisation of the results. Custom views were created, based on a project's needs, to view the crucial information required to support decision making. Fig. 3 shows a typical view for a project where key predicted properties are presented for a set of virtual compounds and the compounds can be visualised docked into the protein structure. The design team uses all these data to help prioritise compounds and aid decision-making, for example prioritising compounds based on favourable protein-docking or superior predicted potency. Final decisions are recorded by flagging de-prioritised compounds as 'not for synthesis'. Ultimately, for synthesised compounds, the measured potency and other data were added (in another custom view). This facilitated analysis and the testing of the original design hypothesis.

One of the more effective additions to the tool was an automatic toxic substructure scan. It is well recognised that the pharmaceutical industry as a whole faces the challenge of predicting toxic outcomes early [23,24]. From various data sources and the wide expertise within AstraZeneca a list of chemical substructures known to be associated with toxic outcomes was collated. As newly designed compounds were added to a design idea they were compared against this list and if a match was found it was flagged up to the user with a brief description of the evidence and some ideas about how to avoid the problem (where known). In this way potentially toxic compounds were flagged up to the design team well in advance of synthesis, enabling informed decisions to be made on progression and possibly avoiding costly attrition later.

Visual planning

The data collected from all the design ideas on a project could be remixed and presented in a variety of ways. One of the most useful views was a visual progress chart. This enabled users to view the 'live' design ideas on the project (i.e. new ideas, those with decisions pending and those that were progressing through to synthesis and testing). There were originally four status categories displayed: new, in design, synthesis/testing and complete. They were shown as four columns of a table. Only the most recently completed ideas were displayed; the remainder (and a fifth category: parked) were archived, although they could be accessed if required. Each design idea was represented by an entry, detailing summary information for the idea: its name, example structure (on mouse hover), priority, number of associated compounds and a hyperlink through to the full description in the wiki. Later the progress chart was expanded to include additional options to colour-code the chart (e.g. by chemical series, by priority or by time spent at the current status level – useful for identifying slow-moving ideas).

Other views were also constructed, including a 'chemists view', to help the laboratory chemists see which compounds had been prioritised for synthesis, and (via a separately written piece of software) a 'testers view', where assay test requests could be made easily for real and virtual compounds – and testers could extract lists of compounds to be plated out for each assay run. Export to other programs for further data visualisation and analysis was also supported.

Tracking design practice

With the use of the Compound Design Database established, it became possible for us to track various aspects of a project's design portfolio and of individual design ideas as they progressed – including the ability to look at averages and trends across and between projects. An improvement activity was then set up to identify suboptimal behaviours and processes, and to implement solutions.

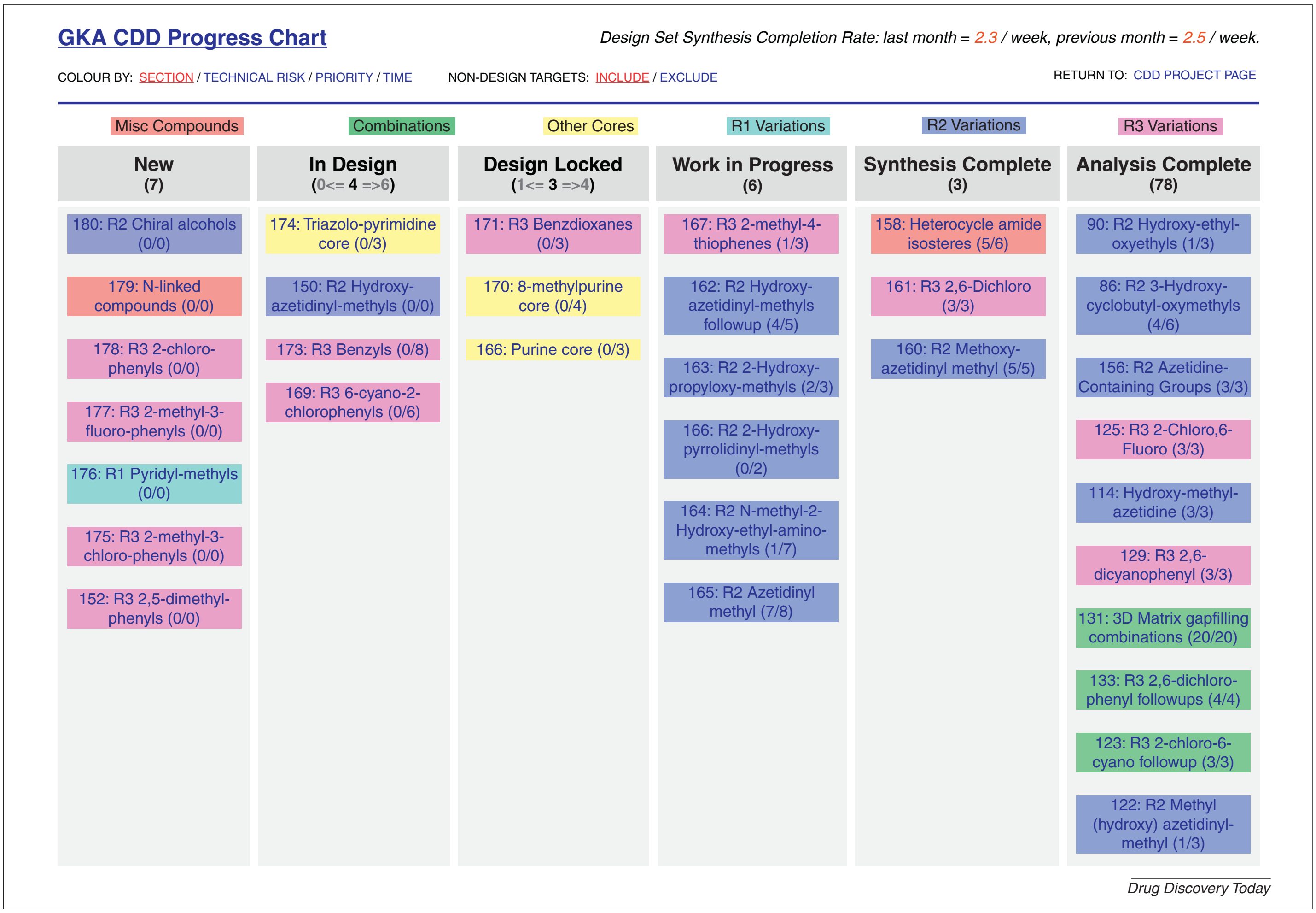
One of the initial changes made was to the visual progress chart. Some additional stages to the process were identified that would trigger actions from team members. The 'design' category was split into two stages: 'in design' and 'design locked' (as shown in Fig. 4); the latter category being a signal to the chemistry team to investigate the synthetic route and order all reagents that could have a long lead time. This successfully eliminated one source of wasted time between design and synthesis. The 'synthesis/testing' category was also split into 'work in progress' and 'synthesis complete'; with the latter being added to enable the tool to track more clearly whether a design hypothesis was in synthesis or in testing. Fig. 4 shows an example progress chart and Table 1 summarises the different status identifiers and the activities associated.

An early finding that led from measuring lead time was that many of our design ideas took much longer than expected to be realised. For one project, half of all design ideas took 27 weeks or more to become actual compounds. This was not a failing of synthetic chemistry but rather it was found that some ideas spent weeks, having been fully worked up by the design team, waiting in a queue before being actioned. It was rationalised that, on these projects, the rate of design was outpacing the rate of synthesis and a backlog was created. This was seen as undesirable because the latest data were not being incorporated into these designs and many designs were never actioned despite a significant design input. Conversely, there were some projects where design ideas were realised quickly. Surprisingly, on further investigation, this was also found to be undesirable because the overall quality of designs going forward was assessed to be lower to keep up with a high rate of synthesis (in terms of knowledge derived from each compound) – with increased pressure on the design team to keep pace with the synthesis rate.

More-efficient design

A mismatch between the capacity in 'make' and the capacity in 'design' is wasteful and undesirable. To face the challenge of a backlog in design the concept of a design inventory was introduced (i.e. a limit was put on the number of different design hypotheses that should be in active design at any one time). The limit for each project was determined as a function of the synthesis rate per design hypothesis for that project, plus some additional capacity to anticipate expected attrition. Thus the limit for 'in design' was set to the current weekly rate of synthesis plus 3.0 (rounded up to the nearest integer) and, similarly, the limit for 'design locked' was set as rate plus 1.5 (rounded up). Effectively, the design ideas were 'funnelled' towards the capacity of the synthesis resource. A lower limit in 'design locked' was also enforced to ensure at least one design hypothesis was fully prepared at all times and ready to be pulled into synthesis. In this way, the rates of design and synthesis were synchronised.

Finally, the imposed limits were exposed in the visual progress chart to make the recommendations clear to all on the project. Visible warnings were presented for projects breaching the limits. So that ideas were not limited and innovation was not stifled, no



GKA CDD Progress Chart

Design Set Synthesis Completion Rate: last month = 2.3 / week, previous month = 2.5 / week.

COLOUR BY: SECTION / TECHNICAL RISK / PRIORITY / TIME NON-DESIGN TARGETS: INCLUDE / EXCLUDE RETURN TO: CDD PROJECT PAGE

Misc Compounds Combinations Other Cores R1 Variations R2 Variations R3 Variations

New (7)	In Design (0<= 4 =>6)	Design Locked (1<= 3 =>4)	Work in Progress (6)	Synthesis Complete (3)	Analysis Complete (78)
180: R2 Chiral alcohols (0/0)	174: Triazolo-pyrimidine core (0/3)	171: R3 Benzdioxanes (0/3)	167: R3 2-methyl-4-thiophenes (1/3)	158: Heterocycle amide isosteres (5/6)	90: R2 Hydroxy-ethyl-oxyethyls (1/3)
179: N-linked compounds (0/0)	150: R2 Hydroxy-azetidinyl-methyls (0/0)	170: 8-methylpurine core (0/4)	162: R2 Hydroxy-azetidinyl-methyls followup (4/5)	161: R3 2,6-Dichloro (3/3)	86: R2 3-Hydroxy-cyclobutyl-oxymethyls (4/6)
178: R3 2-chloro-phenyls (0/0)	173: R3 Benzyls (0/8)	166: Purine core (0/3)	163: R2 2-Hydroxy-propyloxy-methyls (2/3)	160: R2 Methoxy-azetidinyl methyl (5/5)	156: R2 Azetidine-Containing Groups (3/3)
177: R3 2-methyl-3-fluoro-phenyls (0/0)	169: R3 6-cyano-2-chlorophenyls (0/6)		166: R2 2-Hydroxy-pyrrolidinyl-methyls (0/2)		125: R3 2-Chloro,6-Fluoro (3/3)
176: R1 Pyridyl-methyls (0/0)			164: R2 N-methyl-2-Hydroxy-ethyl-amino-methyls (1/7)		114: Hydroxy-methyl-azetidine (3/3)
175: R3 2-methyl-3-chloro-phenyls (0/0)			165: R2 Azetidinyl methyl (7/8)		129: R3 2,6-dicyanophenyl (3/3)
152: R3 2,5-dimethyl-phenyls (0/0)					131: 3D Matrix gapfilling combinations (20/20)
					133: R3 2,6-dichloro-phenyl followups (4/4)
					123: R3 2-chloro-6-cyano followup (3/3)
					122: R2 Methyl (hydroxy) azetidinyl-methyl (1/3)

Drug Discovery Today

FIGURE 4

An example progress chart, built from the design ideas contained in the Compound Design Database. Each coloured box represents a design idea with colouring by user-defined categorisation (in this example by defined substructure types). The numbers that append each design idea name indicate how many compounds have been registered versus the total assigned to that idea.

limits were set for 'new' design ideas, although design activity was limited on these ideas until they had progressed to 'in design'.

The introduction of a design inventory had an enormous impact on the time taken from initial idea to compound in the bottle. Fig. 5 shows three representative projects that were active before and after these changes came into effect. Project 3 is an example of a project where, owing to multistep synthetic routes, the throughput of 'make' was limited and, owing to the rate of design exceeding that of make, a long delay had come into being before design ideas could be progressed into synthesis. It can be clearly seen that the time taken in the latter time period is greatly reduced and so the time from initial idea to compound synthesised is less than half that of the previous situation. The application of this principle not only benefits a 'slow' project but has clear benefits for all projects running at the time. Note that this is not thought to be solely a consequence of the length of time these projects had been running; new projects in the portfolio in the latter time period showed similar mean times for design (data not shown).

The challenge of low quality design, where the rate of synthesis outstrips the rate of design, was also confronted. In simple terms,

TABLE 1

Description of status indicators used in the visual progress chart

Status	Activity
New	Hypothesis stated in outline, design team makes decision to progress based on merits of the hypothesis.
In design	Hypothesis developed, supporting information gathered, potential compounds suggested.
Design locked	Hypothesis and specific compounds approved by design team, synthesis team begin investigation of route and reagents.
Work in progress	Active synthesis in progress.
Synthesis complete	All synthesis completed, testing of compounds proceeding.
Analysis complete	All testing completed, design team have analysed the results and made decisions (including generation of new hypotheses).

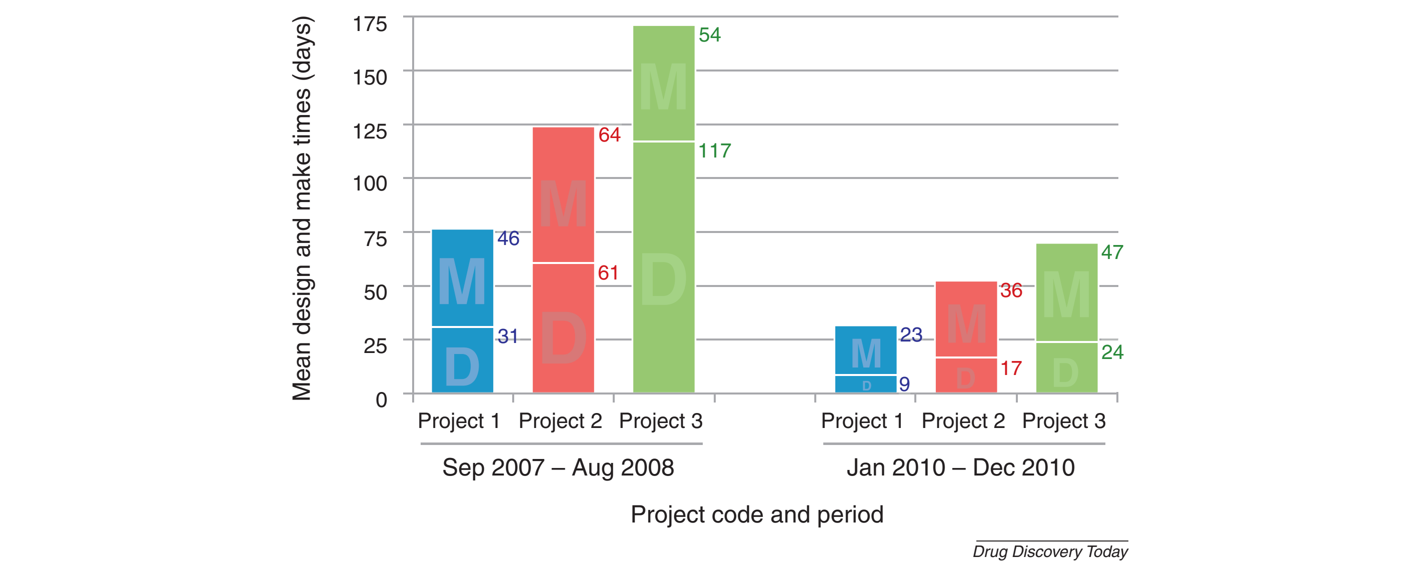


FIGURE 5

The mean number of calendar days for 'design' (D) and 'make' (M) per design hypothesis for three representative projects. Two 12-month periods are shown with the former being before implementation of improvement projects and the latter being after improvements became embedded.

one can imagine reducing the resource in synthesis to equalise the design and make rates. This was deemed to be undesirable on the project at the time. One can also imagine increasing the number of designers on the project to increase the rate of design. The formation of design teams addressed this to an extent, but resource here was still limited. The solution arrived at was to block out sufficient time in the calendar every week solely for the purpose of analysing the latest data and designing new compounds. By protecting this time and preventing meetings from being booked, the existing designers were able to be more productive and thus achieve higher design quality overall.

Reception and legacy

Tools that are perceived to regulate the inherently creative process of drug design can understandably meet resistance from potential users, particularly if they are 'imposed' upon end users. By contrast, the Compound Design Database grew up initially within a small number of projects because they had a need for it. In general, design teams warmly adopted the tool and it continued to evolve in response to these users.

Adoption outside the initial user-group was driven by word of mouth and so the Compound Design Database found positive response from many users. Again the flexibility of the system enabled it to evolve to suit local practices and procedures. Although many people liked it, resistance was met from users who struggled with some of the less user-friendly aspects of the system, in particular the necessity to edit design ideas using wiki markup and the reliance on SMILES for molecular representation. During this phase the use of the Compound Design Database was never mandatory, but it did find use across many projects and multiple R&D sites within AstraZeneca.

Having demonstrated its effectiveness the potential of the tool was clearly recognised. However, the infrastructure underpinning the tool was deemed unsuitable for long-term support. As such, a more robust successor tool was created based on the same philosophy, but using the Django toolkit (http://www.djangoproject.com). The successor tool, named Design Tracker, is significantly more robust than the Compound Design Database and is now standard for use on all drug discovery projects at AstraZeneca and, along with design teams and hypothesis-driven design, the use of Design Tracker is an integral part of how chemistry design operates at AstraZeneca today.

Concluding remarks

Herein the creation of a novel software tool to support collaboration, transparency and data-sharing for compound design in drug discovery is reported. The tool is an essential component of a new way of working, where the emphasis is on hypothesis-driven design through the formation of collaborative design teams.

The software, inspired by wikis and social networks, employs the principles of Web2.0 to create an open-access, highly flexible tool, able to consume data from multiple sources and present slices of these data in meaningful ways to a variety of user roles.

The importance of the visual progress chart for viewing and managing workload and rate of work across disciplines has been emphasised. The use of the tool to track work practises and identify suboptimal parts of the process has also been described.

The concept of a design inventory was introduced to working practices and this has been shown to prevent time wasted at the design stage and to have greatly reduced the time taken to realise design ideas as compounds in the bottle. Similarly the issue of low-quality, under-design was tackled by enshrining time in the schedules solely for the analysis of new data and new design. Together with separate improvements of the 'test' process the time taken for each DMTA cycle to complete has been greatly reduced, enabling more design iterations in the same period of time while simultaneously maintaining or improving the quality of compound design.

Within the lifetime of a project, DMTA represents just one contributory sub-process towards delivering a successful proof of concept and, ultimately, a drug. So far there has been insufficient time to assess whether the overall increase in quality has led to increased project success in development but we feel this is an important step on the journey to achieving this ultimate goal. Moreover, reducing the time spent optimising compounds in discovery is a competitive advantage, regardless of the outcome of clinical studies.

Acknowledgements

The authors wish to acknowledge Alan Birch, Craig Donald, Adrian Gill, Richard Hall, Andrew Leach, Philip MacFaul, William McCoull, Harald Nyman, Charles O'Donnell and Jamie Scott.

References

1 Drew, K.L.M. *et al.* (2012) Size estimation of chemical space: how big is it? *J. Pharm. Pharmacol.* 64, 490–495

2 Horrobin, D.F. (2001) Realism in drug discovery – could Cassandra be right? *Nat. Biotechnol.* 19, 1099–1100

3 Macarron, R. *et al.* (2011) Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* 10, 188–195

4 Ashburn, T.T. and Thor, K.B. (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 3, 673–683

5 Paul, S.M. *et al.* (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* 9, 203–214

6 Plowright, A.T. *et al.* (2012) Hypothesis driven drug design: improving quality and effectiveness of the design-make-test-analyse cycle. *Drug Discov. Today* 17, 56–62

7 Andersson, S. *et al.* (2009) Making medicinal chemistry more effective-application of Lean Sigma to improve processes, speed and quality. *Drug Discov. Today* 14, 598–604

8 Johnstone, C. *et al.* (2011) Creativity, innovation and lean sigma: a controversial combination? *Drug Discov. Today* 16, 50–57

9 Festel, G. (2011) Outsourcing chemical synthesis in the drug discovery process. *Drug Discov. Today* 16, 237–243

10 Subramaniam, S. and Dugar, S. (2012) Outsourcing drug discovery to India and China: from surviving to thriving. *Drug Discov. Today*, http://dx.doi.org/10.1016/j.drudis.2012.04.005 (http://www.drugdiscoverytoday.com).

11 Brodney, M.D. *et al.* (2009) Project-focused activity and knowledge tracker: a unified data analysis, collaboration, and workflow tool for medicinal chemistry Project Teams. *J. Chem. Inf. Model.* 49, 2639–2649

12 Lee, M. *et al.* (2012) DEGAS: sharing and tracking target compound ideas with external collaborators. *J. Chem. Inf. Model.* 52, 278–284

13 Sander, T. *et al.* (2009) OSIRIS, an entirely in-house developed drug discovery informatics system. *J. Chem. Inf. Model.* 49, 232–246

14 Hohman, M. *et al.* (2009) Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery. *Drug Discov. Today* 14, 261–270

15 Williams, A.J. (2008) Internet-based tools for communication and collaboration in chemistry. *Drug Discov. Today* 13, 502–506

16 Mayweg, A. *et al.* (2011) ROCK: the Roche medicinal chemistry knowledge application – design, use and impact. *Drug Discov. Today* 16, 691–696

17 Cheshire, D.R. (2011) How well do medicinal chemists learn from experience? *Drug Discov. Today* 16, 817–821

18 O'Reilly, T. (2007) What is Web2.0: design patterns and business models for the next generation of software. *Commun. Strat.* 65, 17–37

19 Robb, G.R. (2009) Hypothesis-driven drug design using wiki-based collaborative tools. *UK-QSAR and ChemoInformatics Spring Meeting*

20 Rodgers, S.L. *et al.* (2011) Predictivity of simulated ADME AutoQSAR models over time. *Mol. Informatics* 30, 256–266

21 van de Waterbeemd, H. (2009) Improving compound quality through in vitro and in silico physicochemical profiling. *Chem. Biodivers.* 6, 1760–1766

22 Rodgers, S.L. *et al.* (2007) QSAR modeling using automatically updating correction libraries: application to a human plasma protein binding model. *J. Chem. Inf. Model.* 47, 2401–2407

23 Hughes, J.D. *et al.* (2008) Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett.* 18, 4872–4875

24 Leeson, P.D. and Springthorpe, B. (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* 6, 881–890

25 Waring, M.J. *et al.* (2011) Matrix-based multiparameter optimisation of glucokinase activators: the discovery of AZD1092. *MedChemComm* 2, 775–779