DOI: 10.1002/anie.201002238

Small Focused Libraries

Factors Determining the Selection of Organic Reactions by Medicinal Chemists and the Use of These Reactions in Arrays (Small Focused Libraries)

*Tony W. J. Cooper, Ian B. Campbell, and Simon J. F. Macdonald**

compound arrays · drug discovery · medicinal chemistry · small focused libraries · synthetic methods

Synthetic organic reactions are a fundamental enabler of small-molecule drug discovery, and the vast majority of medicinal chemists are initially trained—either at universities or within industry—as synthetic organic chemists. The sheer breadth of synthetic methodology available to the medicinal chemist represents an almost endless source of innovation. But what reactions do medicinal chemists use in drug discovery? And what criteria do they use in selecting synthetic methodology? Why are arrays (small focused libraries) so powerful in the lead-optimization process? In this Minireview, we suggest some answers to these questions and also describe how we have tried to expand the number of robust reactions available to the medicinal chemist.

1. Introduction

Exploitation of the power of organic synthesis for drug discovery is critical for two reasons. First, new drugs are getting rarer. In each of the last three years, less than 25 new medicines were launched, much fewer than half of which were "first-in-class medicines" (i.e. drugs with a novel mechanism of action), despite an estimated combined annual research-and-development expenditure by large pharmaceutical companies of $50 billion.[1] It is estimated that it now costs $1.8 billion to bring a new medicine to the market, and it is forecast that over the next five years, patent expiries will put $209 billion in annual drug sales at risk.[1] It is perhaps unsurprising, therefore, that the pharmaceutical industry is currently undergoing an unprecedented period of upheaval. As a consequence, the whole industry, and medicinal chemists in particular as the primary drug designers, are under relentless pressure to address the extraordinarily high attrition of the clinical candidates they discover: between 93–96% fail to reach the market.[1] In response, medicinal chemists are currently particularly interested in two areas: 1) understanding of structure–toxicity relationships[2] (60% of candidates fail as a result of preclinical toxicity[1]) and 2) control of candidate lipophilicity.[3] Considerable recent literature[4] emphasizes that drug candidates with low lipophilicity (increased polarity) have a reduced risk of attrition. This requirement highlights a challenge faced by medicinal chemists as they seek to capitalize on many of the new synthetic procedures that are published: the compounds used to illustrate these methods frequently have 10 to 10000 times higher lipophilicity than their compound series of interest for drug discovery.

A second reason for exploiting the power of synthesis, however, is the perception that medicinal chemists traditionally use only a tiny fraction of the synthetic transformations available to them.[5] (We make the distinction between reactions used in drug discovery and those used in diversity-oriented synthetic reactions often targeted towards the discovery of tools for chemical biology.)[6] Whole reaction classes appear to be almost ignored and enantioselective methods for the synthesis of structures with multiple stereogenic centers generally eschewed. It seems reasonable to presume that if a wider range of reactions were used, more physicochemically diverse compounds could be made. The use of a wider range of reactions would provide more options for the introduction of polarity into target compounds and would lead to the generation of compounds with increased novelty for patenting purposes (as companies in the pharmaceutical industry frequently work on very similar series).

This highly selective use of methodology is particularly apparent in the reactions that medicinal chemists use when

[*] T. W. J. Cooper, I. B. Campbell, Dr. S. J. F. Macdonald
Respiratory CEDD, GlaxoSmithKline
Gunnels Wood Road, Stevenage, SG1 2NY (UK)
Fax: (+44) 1438-762-302
E-mail: simon.jf.macdonald@gsk.com

they make arrays or small focused libraries (we use the term "arrays" henceforth). Arrays are now a powerful tool for medicinal chemists. They emerged from the combinatorial chemistry era and enable the rapid simultaneous synthesis and purification of larger numbers of compounds by the same reaction (typically 12–96 compounds). Automated procedures for liquid handling, purification, and analysis are widely used. Typically, in an array, one part of the molecule is held constant, while another part is varied. The reagents used to introduce the varying part of the molecule are often referred to as monomers or building blocks (we use the term "monomers" henceforth).

The purpose of an array is usually to establish structure–activity relationships (SARs) against a biological target. Data from the biological testing of arrays, and particularly two-dimensional arrays, in which two parts of the structure are varied simultaneously, provide "information-rich" SARs, which are much more valuable than SARs from individual compounds made sequentially (as described in Section 3.2).

Given this background—the need to reduce attrition, to make compounds with lower lipophilicity, and to use a wider range of synthetic methodology (particularly in the preparation of arrays)—we recently investigated what reactions were being used by lead-optimization medicinal chemists in respiratory drug discovery. The consequences and implications of this investigation are described herein.

In Section 2 of this Minireview we describe 1) the data that we collected on the range of reactions we used; 2) how we sought to expand the reactions we used; and 3) the criteria that the medicinal chemist tends to use in reaction selection. An understanding of these criteria may be helpful to synthetic organic chemists who are interested in developing methodology for use by medicinal chemists.

In Section 3 of the Minireview, we describe the value of arrays to medicinal chemists on the basis of our experience. We outline the advantages we have found and comment on questions that are often asked in relation to the use of arrays in medicinal chemistry.

Tony Cooper has over 25 years' experience in the pharmaceutical industry as a medicinal chemist at GlaxoSmithKline. He has previously worked in combinatorial chemistry and is now the Group Leader responsible for the Compound Arrays Team delivering array support to lead-optimization programs within the Respiratory Centre of Excellence for Drug Discovery in Stevenage (UK).

Ian Campbell has almost 30 years' experience in the pharmaceutical industry, at GlaxoSmithKline. He has previously managed teams within medicinal chemistry and combinatorial chemistry and is now a Group Leader responsible for a synthetic chemistry group who provide chemical, technical, and outsourcing support to the Respiratory Centre of Excellence for Drug Discovery in Stevenage (UK).

Simon Macdonald has over 20 years' experience as a medicinal chemist in the pharmaceutical industry and has spent his entire career at GlaxoSmithKline in its various incarnations. He is currently a director of medicinal chemistry in the Respiratory Centre of Excellence for Drug Discovery in Stevenage (UK).

2. Expanding the Range of Reactions Useful for Medicinal and Array Chemistry

2.1. What Reactions Are Used in Medicinal Chemistry...?

Within GlaxoSmithKline (GSK), the discovery of clinical candidates by lead optimization is carried out by the Centres of Excellence for Drug Discovery (CEDDs), which are organized according to therapeutic area. Each CEDD is a semiautonomous unit with its own processes and culture. The data and analysis presented herein are historical data (primarily from lead-optimization activities) from the Respiratory CEDD, which is based in Stevenage in the UK and Upper Merion in the USA. As the name implies, its focus is on the discovery of clinical candidates for respiratory diseases. There are around 100 chemists in the CEDD, which incorporates a centralized arrays team.

In 2005, we initiated the debate about the reactions used by the chemists in the Respiratory CEDD. There was a widespread perception that the range of reactions was quite narrow. To investigate this matter, we surveyed 4800 contemporaneous reactions carried out by the chemists in the CEDD. Reactions were grouped into 22 general classes; multiple reactions as part of an array were counted as one reaction (Figure 1, top) and separately (Figure 1, bottom).

The data only provide a crude snapshot of the breadth of chemistry used. Furthermore, we were aware that in the course of a lead-optimization program, the same types of reactions tend to be utilized over and over again, which may skew the snapshot. There are also notable gaps: no reductive aminations were carried out over the period of the survey, although in our experience, these transformations are traditional work-horse reactions in medicinal chemistry. Nonetheless, 63% of the 4800 reactions fell into four reaction classes (alkylations, condensations (amides and sulfonamides), palladium-catalyzed couplings, and protecting-group manipulations), and 89% of the reactions fell into just 10 reaction classes (the four already mentioned plus halogenations, heterocycle formation, hydrolyses, metalations, oxidations, and reductions). Some reaction classes rarely featured (rearrangements, radical reactions, metatheses, and cycloadditions), and reactions that generate new stereocenters were noticeable by their absence (see Section 2.2). Of 577 array reactions, 87% fell into just three reaction classes: alkylations, condensations, and palladium-catalyzed couplings (Figure 1, lower pie chart).

2.2. ...and Why?

What factors lead to the use of this limited range of reactions by medicinal chemists? Following the survey and extensive in-house discussions, the following emerged as likely reasons:

1) Owing to commercial and competitive pressures, the aim of medicinal chemists is to discover a clinical candidate as fast as possible, so robust, established methodologies are preferred. Because most of their synthetic targets are "low-value" targets,[7] they have little time to develop or optimize a

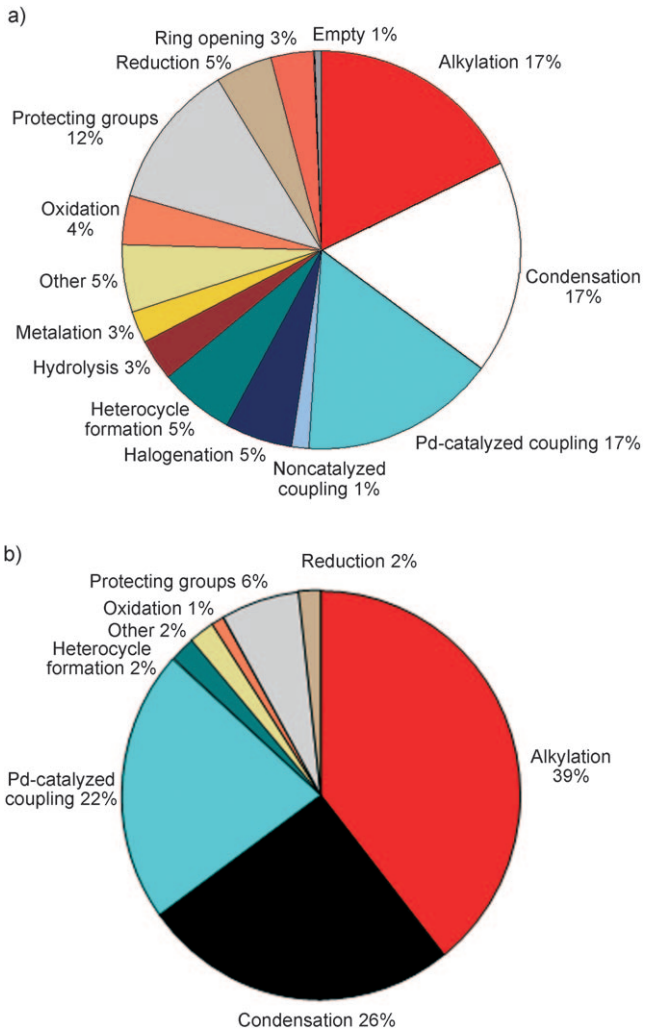

Figure 1. a) Pie chart showing the main categories of the 4800 reactions studied in the survey. Arrays were counted as single reactions. b) Pie chart in which the 577 array reactions have been separated out and divided into the main reaction categories.

reaction or route. Any reaction optimization tends to occur only as the value of the target starts to rise, typically as potential clinical candidates begin to be identified.

2) A corollary to the focus of medicinal chemists on the discovery of clinical candidates is that they may be reluctant to use a broader range of reactions if they are generally deemed unnecessary. Their primary focus is on drug discovery with synthesis as a tool; as a result, maintaining an awareness of the latest synthetic methodology may not be considered a priority.

3) Most drug molecules incorporate a variety of functional groups, such as heterocycles, amides, alcohols, amines, and carboxylic acids. The synthetic methodology used must be tolerant of this functionality, ideally without the need for elaborate protection strategies. When new synthetic methodology is published with limited examples of its use, and when the examples presented do not demonstrate functional-group tolerance, uptake by the medicinal chemist is usually slow. In contrast, when wide exemplification is available—for example, in the case of the Buchwald–Hartwig amination, Suzuki and Stille cross-coupling reactions, and Sharpless click chemistry—uptake is much faster and wider.

4) Physicochemically, ideal druglike compounds are characterized by low lipophilicity (clog$P < 3$) and high solubility in aqueous media.[4] The introduction of these properties into a molecule is often a challenge, particularly while maintaining the pharmacological profile of the compound. For example, the impact of the addition of different fragments to a core on the overall lipophilicity of a compound is shown in Figure 2. The difference in the lipophilicity of the product following the addition of a substituted *N*-pyridyl or an *N*-aminophenyl to a template is significant (approaching 1000-fold in the example). Furthermore, the introduction of polar fragments or fragments containing functionality can sometimes reduce the yield of the reaction or the methodology used. When new methodology is published and the examples presented are highly lipophilic, uptake by the medicinal chemist is usually slow. For example, in one reported array demonstrating methodology for the medicinal chemist, most of the products possessed high lipophilicity (clog$P > 4$; Figure 3).

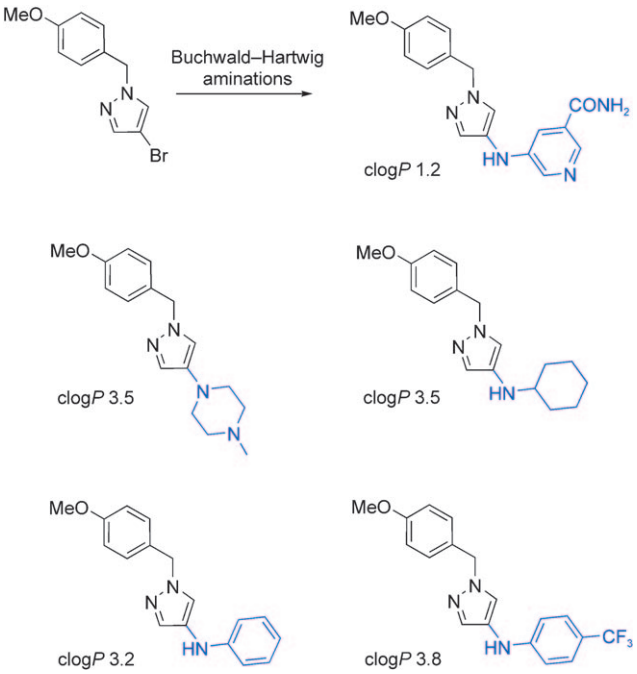

Figure 2. Hypothetical small array in which structural fragments (in blue) are coupled to a pyrazole. Note in particular 1) the range of functionality introduced, particularly in the first two analogues; 2) the impact of the introduced fragment on the lipophilicity of the overall molecule; 3) the impact of the fragment on the aqueous solubility of the overall molecule (the second analogue is likely to be much more soluble than the fifth).

5) Reactions that generate new stereocenters (and particularly those which are not stereoselective) have historically been used less frequently, probably owing to the extra work required both in the drug-discovery process and downstream in drug development.[8] Issues that might arise include the

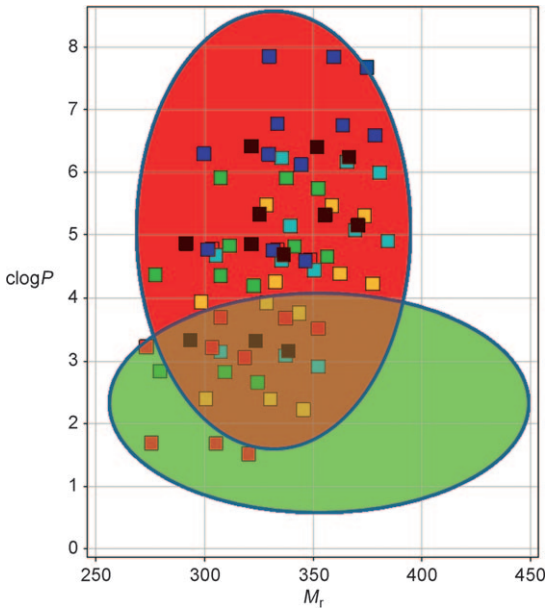

Figure 3. Plot of calculated lipophilicity (clog*P*; calculated with the software Daylight v481) against molecular weight. Each colored square represents a member of an array from the literature. The red shaded area is "traditional-chemistry" space. The green shaded area is the "druglike" space in which the medicinal chemist seeks to operate.

determination of the biological activity of the other isomer. How stable is the stereocenter under metabolizing conditions in vivo? How can the desired isomer be purified easily on a large scale? In a recent lead-optimization program involving compounds with two stereocenters, a full-time employee was dedicated solely to the separation of isomers: an expense that is not readily justified, even in the pharmaceutical industry. However, clinical candidates containing stereocenters may be associated with lower attrition in development, as highlighted by a recent report.[9] When combined with the power of modern stereocontrolled reactions, these factors suggest that the inclusion of stereocenters in drug molecules may increase.

Table 1: Some structures and functionalities that medicinal chemists prefer to avoid in their final compounds[10] (unless the reactivity is intrinsic to the mechanism by which they operate).[a]

Class	Examples
reactive functionality	aldehydes, epoxides, imides, 1,2-dicarbonyl compounds, Michael acceptors, alkyl halides, acyl halides, sulfonyl halides, esters
N–O and N–N bonds (unless part of a heteroaromatic compound)	oximes, hydroxylamines, hydrazones, hydrazides, nitrones, azo and nitro groups
compounds with redox potential	quinones, dihydroaromatic compounds
miscellaneous	anilines, fluorescent groups, rhodanines, furans, pyrroles

[a] Many drugs do actually feature some of these structures and functionalities,[10] and each medicinal chemist and company will have their own list and prejudices.

6) The reagents, catalysts, and monomers for the transformation should be commercially available; otherwise, the methodology is less likely to be used.

7) Medicinal chemists are under considerable pressure to deliver high-quality candidates that will survive the development process to become marketed drugs (see Section 1).[4] They are bombarded with information[10] relating to structural types (furans, pyrroles) or functionality (anilines, nitro groups) that represent an increased risk of attrition in the development process. This is not to say that there are no drugs which contain such features (there are); rather, if these features are avoided, the risk of attrition may be reduced. As a consequence, reactions that are only exemplified with such templates or functionality are less likely to be used (Table 1). Conversely, given that many drugs contain heterocycles, basic centers, and acidic NH groups, reactions are more likely to be used by medicinal chemists when examples that include such features are presented.

2.3. Practical Criteria for a Robust New Reaction for an Array

In the context of the preparation of arrays of compounds, the above Section 2.2 points are even more important. However, additional factors also need to be considered. For single, hand-crafted reactions, there are few critical practical considerations, but when working with multiple compounds simultaneously, it is unrealistic for the medicinal chemist to develop bespoke procedures for each reaction in the array. For arrays—where compromises are inevitable—the practical criteria for the ideal reaction include the following:

1) The reaction should not require extremes of temperature or precise reaction times.
2) Reagents/catalysts are ideally tolerant of oxygen or atmospheric moisture.
3) A variety of reaction concentrations and the use of excess reagents should be tolerated.
4) The reaction should be compatible with the use of polar solvents: many of the substrates/monomers used by medicinal chemists are insoluble in nonpolar solvents.
5) The reaction should be compatible with a variety of solvents or even mixed solvents: sometimes the substrates/monomers will dissolve only in one solvent.
6) The reaction should allow the flexibility to mix the monomers with the catalyst/reagent for ease of dispensing.

Other practical considerations include:

1) the importance of simple workup and purification procedures: these steps typically take 5–10 times as long to carry out as the actual synthesis of an array
2) the flexibility to dispense the reagent as a slurry (this way, the amount of solid per unit volume is known) if required
3) the minimization of weighing operations.

Overall, for any new methodology to be useful for arrays, the key is to determine a range of conditions that will tolerate a wide variety of polar functionalized monomers.

2.4. Enlisting the Help of Academia to Broaden the Range of Reactions

As discussed in the preceding sections, there are valid reasons for the use of a limited range of synthetic reactions by medicinal chemists. However, a wider variety of reactions is still desirable. This section outlines how we are seeking to address this issue.

The overriding factor for the lack of diversity of reactions used by medicinal chemists is the lack of time these chemists have available to explore or optimize alternative methodology. The more complex the synthesis of a target or the more development that synthesis requires, the greater the justification required for that particular target.[11] As mentioned in Section 2.2, when most of the targets are "low-value", this justification is not easy. To expand the range of robust reactions with tolerance of a wide range of functionality, a significant additional resource would be required. As this resource was not readily available in-house, we explored alternative options. One of these options was to cofund the development of robust synthetic methodology for medicinal chemistry with the Engineering and Physical Sciences Research Council (EPSRC), the leading funding agency of the UK Government. The EPSRC supports much of the academic research and training in chemistry in the UK.[12] A major advantage of this approach is that it capitalized on the outstanding synthetic expertise of chemistry academics in the UK.

As a consequence, the EPSRC, in partnership with the Respiratory CEDD, published two calls to UK academics for research proposals entitled "Reactions for Array Chemistry: Reactions, Design and Interpretation" in 2006 and 2007. The total value of both calls was over £4 million, with 12 collaborative projects and approximately 20 PhD and postdoctoral positions supported at universities in the UK. A unique aspect of this scheme is that up to 50% of the period of appointment to the PhD and postdoctoral positions is spent at the GSK Medicines Research Centre in Stevenage in the UK. One benefit of this arrangement has been the opportunity to provide the PhD students and postdoctoral research assistants with additional industry-based training while also exposing them to the practices, culture, and chemistry of drug discovery in a big pharmaceutical company.[13] Numerous studies have already been published, and by this benchmark, the calls have been very successful.[14] However, one challenge has been to marry the pure research perspective of academia with the applied research expertise of industry to genuinely add new robust reactions to those available to medicinal chemists.[15]

The advantages of arrays for medicinal chemistry and the desirability of reactions for array chemistry are discussed in Section 3, but in some way these calls for research proposals mirror an initiative funded by the National Institutes of Health in the USA entitled "Chemical Methodologies and Library Development". The aim of this initiative is to "develop efficient, general, state-of-the-art methodologies for the design, synthesis, analysis, and handling of chemical diversity libraries. The goals of the program are to discover and to implement fundamental new chemistry that will facilitate access to high-quality libraries of expanded diversity."[16] This initiative has led to the creation of five such research centers run by well-known principal investigators.

3. The Value of Arrays in Medicinal Chemistry

3.1. The Use of Arrays in the Respiratory CEDD GSK

Arrays are widely used as a powerful tool for lead discovery and optimization in the Respiratory CEDD and are predominantly synthesized by a dedicated arrays team. The Respiratory CEDD array team is typically composed of three or four full-time employees with one or two university students on a one-year placement in industry. Owing to their focus on and expertise with arrays, we have found that the array chemists can make, purify, analyze, and register array compounds more efficiently than the program medicinal chemists. They are frequently also more effective in delivering a higher percentage of products from the array in greater yield and purity.

The team has a unique skill set and mindset. We have found that an array chemist should be highly organized, show attention to detail, be manually dexterous, be comfortable with repeatedly delivering to deadlines, and have an ability to work with often introverted and occasionally obstreperous program chemists! This combination of characteristics is uncommon amongst chemists.

In a typical scenario, the lead-optimization-program team (charged with delivering the clinical candidate) liaises with the CEDD array team about the synthesis of an array. The program team then prepares the bespoke monomers and substrates.[17] These compounds are handed over to the array team, who order the remaining monomers, synthesize and purify the array, generate the characterizing analytical data, register the compounds in the corporate database, and dispense and dispatch the samples for testing. The vast majority of purifications are carried out by using automated mass-directed (fractions are collected once the desired molecular weight is detected by in-line mass spectrometry) reversed-phase HPLC equipment. For 80% of arrays, it takes approximately two weeks from the initiation of synthesis to the delivery of the compounds for testing. The arrays team also plays a significant role in the continual enhancement of array design and the development of new data-visualization techniques; these advances are then shared with the program teams.

From data gathered over a five-year period, the CEDD synthesized and submitted around 8500 compounds for biological screening on average each year. Remarkably consistently, in each six-month time period, around 66% of these compounds were synthesized in an array format.

During this five-year period, each lead-optimization-program team within the CEDD had the flexibility to use the array-team resource as they saw fit.[18] In other words, the program medicinal chemists followed their own philosophy as to the use of arrays and who made them—there was no management directive imposed. The use of arrays as part of the program strategy was driven in part by chemical tractability and in part by the preferences of program

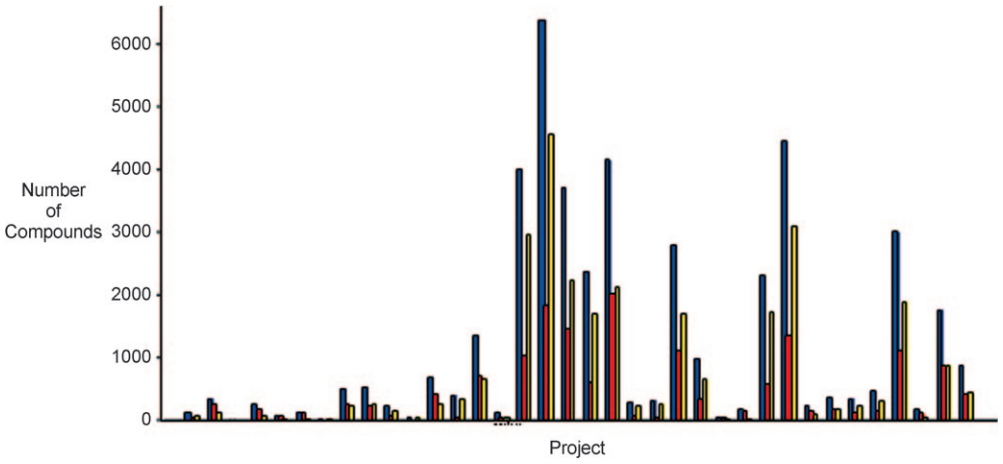

Figure 4. Graph showing the number of compounds made in individual projects by the Respiratory CEDD at GSK during a five-year period. In each case, the total number of compounds (blue column) is then split into those made as singletons (red column) and in arrays (yellow column). Each blue column represents the sum of the corresponding red and yellow columns.

chemists (Figure 4). For example, in 2007, two out of nine program teams fully integrated arrays into their lead-optimization program (in which they were central to the determintaion of SARs), a further three of the nine showed interest, and the remaining four only occasionally requested arrays when asked to do so (for the generation of "nice-to-have data"). However, we have observed that once program-team chemists become comfortable with using the array team, they remain high users. Once a level of trust and a working relationship have been established, the partnership is seen not only as adding value, but the use of arrays as part of the lead-optimization process is seen to offer advantages (see Section 3.2). The system works best when there is a two-way discussion between the program team and the array team, with each team regarding the other as their peers.

3.2. Advantages of Arrays

Within the Respiratory CEDD, the design of each array is predominantly the responsibility of the program chemist, although the best arrays are usually designed in conjunction with the arrays team (and often computational chemists).[19] Considerable emphasis is placed on the incorporation of a high degree of design into the members of each array and the array itself. The design generally involves in silico profiling of the array prior to synthesis, and several "design iterations" are sometimes required before the final version of the array is determined.

Our experience is that the use of arrays in an iterative lead-optimization (or lead-discovery) process has numerous benefits:

Arrays facilitate a rapid and comprehensive exploration of SARs. Two-dimensional or "squared" arrays (in which two parts of the compound are varied simultaneously) are particularly rich in SAR information and can provide clear guidance on the way forward. The risk associated with medicinal-chemistry decisions based on biological data from arrays of compounds is lower than that of decisions based on single or very few compounds.

As an example, a hypothetical synthetic scheme for the preparation of a squared array of compounds is shown in Figure 5 with a heat map indicating their potencies. Compounds represented by dark green are more potent than those represented by light green; yellow indicates compounds that are inactive. Gray indicates that the synthesis of the compound failed. It is clear that the *p*-cyanobenzamides are the most potent compounds, except when the amide substituent is R^4 or R^{12}. *m*-Cyanobenzamides are also active, although less so, again except when the amide substituent is R^4. If only the R^4- or R^{12}-substituted *p*-cyanobenzamides had been made as single compounds, a rich seam of activity might have been missed. The *o*-cyanobenzamides are inactive except for the R^4- and R^{12}-substituted amides. From the data for the whole array, it is now easy to see that the potency of the R^4-substituted *o*-cyanobenzamide is noticeably different to the potency of related analogues. Thus, arrays, and squared arrays

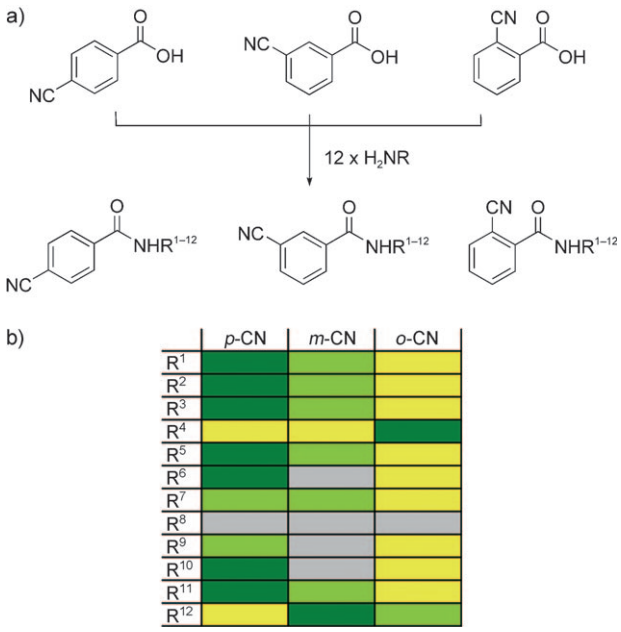

Figure 5. a) Hypothetical 3×12 two-dimensional array in which the activity of an *ortho*, *meta*, or *para* cyano substituent is determined with 12 different amides. b) Heat map showing the potency of compounds in the array. Dark green indicates higher potency, light green indicates lower potency, and yellow indicates that the compound is inactive. Gray cells indicate that the synthesis of the analogue failed.

in particular, can reduce the risk of missing important leads or being deceived by the anomalous activity of one compound. A much higher quality decision can often be made about what targets should be synthesized in the next iteration.

Anomalous data. On viewing the data for the R^4- and R^{12}-substituted *o*-cyanobenzamides and the R^{12}-substituted *p*-cyanobenzamide in the light of the data from the rest of the array, the experienced medicinal chemist would probably regard these data points as "anomalous" and almost certainly have the compounds retested (and perhaps repurified and retested) to confirm or disprove the data. If the data are confirmed on retesting, then perhaps these compounds represent new strands of the SAR that might have been missed if a squared array of analogues had not been prepared.

Thus, when one compound shows unexpected activity and it is part of an array of data, it is much easier (and almost certainly more accurate) to predict whether the data point reflects a genuine difference in SAR or whether it is a faulty data point. In other words, the activities of the remainder of the compounds in the array act as "internal standards" in the data set. In our experience, such data anomalies are common. When these anomalies are seen in the context of an array of data, better decisions can be made than when compounds are made individually or in very small sets. In a real-life example from a phosphodiesterase 4 inhibition program, the effect of a chiral center was unclear from initial testing of a single analogue, but across an array of analogues, the preferred configuration became apparent.

Identification of seams of activity. The synthesis of six members of the array failed (Figure 5). From the data for the whole array, a higher quality decision can be made as to whether a second attempt at their synthesis should be made. In this case, a second attempted synthesis would perhaps only be justified for the R^8-substituted *p*-cyano analogue.

Deciding whether to reattempt synthesis of failed analogues. It is common in the lead-optimization phase that multiple biological and/or physicochemical parameters need to be optimized in parallel. In our hypothetical example, it might be best to avoid trying to optimize the R^4- or R^{12}-substituted *o*-cyanobenzamides, because it would be very easy to lose the potency in this series while trying to optimize other parameters. The *p*- or *m*-cyanobenzamides, on the other hand, are likely to maintain their potency while other parameters are optimized.

Increased chances of serendipitous advances. Despite major improvements in compound design through the use of in silico tools, many significant advances in the lead-optimization process result from chemists following a hunch or a speculative hypothesis. These ideas often break the established SAR, the received wisdom, or the prejudice of the program. Frequently, the synthesis of such compounds is not publicized by the chemist until they have been tested, because the program leader may deem them unnecessary. In contrast, the preparation of arrays provides room for such hunch or speculative compounds to be made efficiently (and openly!). In fact, our best practice in designing arrays is for 10–20% of the monomers to be deliberately selected to follow a hunch or break the SAR rules.

Rapid generation of a "back catalogue". Finally, array-driven lead optimization leads to an excellent back catalogue of compounds much more rapidly than the synthesis of single compounds or very small arrays. If the aims of the program should change (the envisaged route of administration, for example) or if a certain parameter suddenly becomes important (for example, selectivity against p450 isoforms), there is a rich set of compounds of reasonable diversity already available. These compounds can be immediately accessed for SAR analysis and/or further screening, which can save considerable time. As the back catalogue gets larger, the chance that the solution to a new issue already lies within the database of structures and results becomes greater.

3.3. Common Questions Relating to the Use of Arrays in Medicinal Chemistry

In this section, we answer common questions about the use of arrays in medicinal chemistry on the basis of our experience.

How long do larger arrays take to make and purify? A frequent comment about the synthesis of larger arrays ($24+$) is that the synthesis and purification of single compounds or very small arrays (<6) is much quicker. In our arrays team, 80% of 24-membered arrays are typically synthesized, purified, characterized, registered, and formatted for screening within 2 weeks. The process is aided by automated platforms. Also, biological screening within GSK is heavily automated, with assays running weekly. So although single compounds or small arrays may be prepared in less than 2 weeks, the time saving is often not huge. Furthermore, the data available from an array of compounds within a 2–3 week period may be of greater value than data available from single compounds or small arrays in 1–2 weeks.

How much of the target compound is made in the array? In most of our arrays, at least 0.1 mmol of the substrate is used for each final product (with a molecular weight of 350–550) in the array. Even with moderate yields, purification by automated mass-directed reversed-phase HPLC can deliver 10–20 mg of the final compound. In GSK, this is sufficient sample for in vitro testing in multiple assays and preliminary pharmacokinetic assessment before resynthesis is required.

Does testing arrays swamp the available screening resource? If the initial assay in the screening cascade has very low throughput, then the preparation of arrays is clearly inappropriate. In our experience in GSK, however, this situation has rarely been encountered, and when it has, appropriate triaging of compounds has been effective. (We recognize that big pharmaceutical companies are able to allocate larger resources to screening than small companies or academic groups.)

Is it not a more effective drug-discovery strategy to design and synthesize fewer compounds than to make arrays? Our view is that highly designed arrays of compounds represent the best of both worlds. As mentioned earlier, arrays also facilitate serendipitous discoveries and allow scientists to follow individual hypotheses without consuming excessive resources.

Does the application of Free–Wilson principles obviate the need for arrays? Free–Wilson[20] quantitative structure–activity relationships can be used to predict the activity of compounds in an array following the synthesis and testing of a few selected members. This technique is often powerful. However, Free–Wilson analysis makes the assumption that one part of the structure does not affect the binding or conformation of another part of the structure; this assumption frequently does not apply. As Breitenbucher and co-workers[21] point out: "…despite this knowledge, many medicinal chemistry papers continue to report uni-dimensional analoging. In these cases, chemists seem prepared to accept the risk of missing key pieces of SAR due to non-additive behavior, or are perhaps unaware of the limitations of the additive assumption." Thus, in cases in which Free–Wilson principles cannot be applied, arrays may be of great value.[22]

How long does it take to analyze the data from large numbers of compounds? Without computational visualization tools (Spotfire, for example), data analysis would indeed be time-consuming. However, in our own experience, Spotfire visualization and associated property calculations are extremely powerful SAR tools.

What if there is insufficient intermediate available to make an array? This situation occurs frequently! If intractable chemistry limits the quantity of intermediate available, then the opportunities for preparing arrays are severely constrained.[23]

Is it not more efficient for the medicinal chemists to prepare the arrays? In our experience, the synthesis of arrays by expert array chemists liberates more time for expert medicinal chemists to do medicinal chemistry. We have found consistently that an expert array chemist is more effective in delivering more products from an array in higher yield and purity than the "jack-of-all-trades" medicinal chemist.

4. Outlook

We have described herein how within the Respiratory CEDD in GSK, we have tried to expand the range of reactions that our medicinal chemists use and particularly those used in the preparation of arrays. It is a "work in progress" with much yet to be achieved towards our goal of ultimately delivering compounds with increased novelty, lower lipophilicity, and an increased chance of surviving the development process. We have outlined some of the properties a new reaction should have to be of widespread use to medicinal chemists. We have also detailed the benefits of arrays according to our drug-discovery experience.

We thank Robin Carr and Ian Churcher for critical reading of this manuscript and invaluable suggestions. We thank Chris Edlin and Heather Hobbs for the collation and analysis of the reactions carried out in the Respiratory CEDD. We particularly thank our collaborators at the EPSRC—John Baird, Emma Jones, and Zoe Brown—without whose help, advice, and effort, the "New Array Chemistry Calls" would not have been possible. We also thank and acknowledge our internal GSK partners who have provided advice and support: Malcolm Skingle, Tamsin Sayer, Nicky Pattrick, and Jane Lewis in Academic Liaison, and Iain McLay in Molecular Discovery Research. Within the Respiratory CEDD, we received considerable input and counsel from many individuals, but particularly from Dave Allen, Peter North, and George Hardy. We thank past and present members of the array team: Heather Barnett, Heather Hobbs, and Natalie Wellaway. We thank the industrial supervisors, the academic supervisors, and the students themselves—around 50 people—who have committed and are committing their time and energy to "New Array Chemistry" research. Finally, we are grateful for funding by the EPSRC, which allowed the two calls to go ahead and is supporting 12 industry–academic collaborations.

Received: April 15, 2010
Published online: September 21, 2010

[1] a) S. T. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg, A. L. Schacht, *Nat. Rev. Drug Discovery* **2010**, *9*, 203–214; b) I. Kola, J. Landis, *Nat. Rev. Drug Discovery* **2004**, *3*, 711–715; c) H. Kubiyi, *Nat. Rev. Drug Discovery* **2003**, *2*, 665–668; d) an analysis reported in a KMR Pharmaceutical Benchmarking Forum R&D General Metrics Study Presentation on May 4, 2006 stated that the overall new-medical-entity-attrition median (i.e. the median percentage of drug candidates that didn't make it to market) for 2002–2006 in the pharmaceutical industry was 91%.

[2] J. A. Kramer, J. E. Sagartz, D. L. Morris, *Nat. Rev. Drug Discovery* **2007**, *6*, 636–649.

[3] P. D. Leeson, B. Springthorpe, *Nat. Rev. Drug Discovery* **2007**, *6*, 881–890.

[4] Log *P* is the partition coefficient of a compound between octanol and water. Lipophilicity can be calculated and described as clog *P* (with varying degrees of accuracy), which is a very widely used parameter in medicinal chemistry. ACD labs offer a free clog *P* calculator http://www.acdlabs.com/download/logp.html. In a seminal study, Lipinski et al. investigated the lipophilicity of oral drugs (among other compounds) and suggested that for good absorption and permeability (important factors in the discovery of oral drugs) the clog*P* value should be less than 5 (C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25). In fact, this value is now viewed as being at the upper end of desirable lipophilicity for an oral drug; clinical compounds with a clog*P* value between 1 and 3 are preferred. Correlations between the high lipophilicity of clinical candidates and increased attrition have become a recurring theme. Leeson and Springthorpe[3] from AstraZeneca commented on drug promiscuity (the tendency of a compound to modulate biological processes other than the desired process): "The consequences … of the marked increase in lipophilicity include a greater likelihood of lack of selectivity and attrition in drug development." Researchers from Pfizer wrote of "an increased likelihood of toxic events [that] was found for less polar, more lipophilic compounds" (J. D. Hughes, J. Blagg, D. A. Price, S. Bailey, G. A. DeCrescenzo, R. V. Devraj, E. Ellsworth, Y. M. Fobian, M. E. Gibbs, R. W. Gilles, N. Greene, E. Huang, T. Krieger-Burke, J. Loesel, T. Wager, L. Whiteley, Y. Zhang, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4872). Gleeson from GlaxoSmithKline noted "the need to focus on a lower molecular weight and logP area of physicochemical property space" (M. P. Gleeson, *J. Med. Chem.* **2008**, *51*, 817–834). In relation to compound permeability (the ability of a compound to cross membranes): "log*D* and molecular weight are the most important factors" (M. J. Waring, *Bioorg. Med. Chem. Lett.* **2009**, *19*,

2844 – 2851). We found that "adding more lipophilicity by adding more aromatic rings … is likely to increase the risk of attrition" (T. J. Ritchie, S. J. F. Macdonald, *Drug Discovery Today* **2009**, *14*, 1011 – 1020). Regarding compounds for lead optimization: "key properties [such as lipophilicity] of recently developed clinical candidates and advanced lead compounds have been shown to differ significantly from those of historical leads and drugs" (G. M. Keserü, G. M. Makara, *Nat. Rev. Drug Discovery* **2009**, *8*, 203 – 212).

[5] For an analysis of reactions used in the large-scale preparation of drug-candidate molecules, see: J. S. Carey, D. Laffan, C. Thomson, M. T. Williams, *Org. Biomol. Chem.* **2006**, *4*, 2337 – 2347.

[6] Diversity-oriented synthesis (DOS; T. E. Nielsen, S. L. Schreiber, *Angew. Chem.* **2008**, *120*, 52 – 61; *Angew. Chem. Int. Ed.* **2008**, *47*, 48 – 56) provides valuable tools for chemical biology. However, many of the molecules generated are structurally complex and have relatively high molecular weights and lipophilicity. These properties hinder their use as potential leads for oral-drug-discovery programs. High structural complexity reduces hit rates and can make optimization chemistry difficult (M. M. Hann, A. R. Leach, G. Harper, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856 – 864).

[7] As a compound gets closer to becoming a clinical candidate, it becomes a "higher-value" compound, and more time is often invested in refining the synthetic route.

[8] Clearly, there are many exceptions to this statement. However, the introduction of a stereogenic center (or a further stereogenic center) in a lead series is rarely undertaken lightly, particularly when the stereogenic center is not embedded in a commercially available molecule or cannot be introduced with good stereoselectivity. "Development" refers to the activities between the discovery of the clinical candidate and its launch onto the market. It is a highly resource intensive and expensive process and includes scaling up, toxicity studies, clinical trials, and registering the drug with the regulatory authorities (e.g., FDA) regulatory. When a candidate has numerous isomers, there can be a significantly increased burden to show that the biological activity resides in a particular isomer and that the properties and quantities of the undesired isomers are both known and controlled.

[9] F. Lovering, J. Bikker, C. Humblet, *J. Med. Chem.* **2009**, *52*, 6752 – 6756. In questioning whether "more is lost … than gained by avoiding cutting edge synthetic technologies … [and not] taking a little more time to introduce chiral centers", one of the referees raised a fair point. In our experience, considerable nerve is required from the medicinal chemist to justify the "time delays" that might result.

[10] Frequently, filters are used to flag or exclude particular types of structure or functionality in an attempt to decrease drug-discovery attrition: J. B. Baell, G. A. Holloway, *J. Med. Chem.* **2010**, 49, 1433 – 1441 and references therein; see also: http://ncgc.nih.gov/projects/cruzain/Cruzain_qHTS_Supplemental_Table_Exclusion_Filters.xls.

[11] S. J. F. Macdonald, P. W. Smith, *Drug Discovery Today* **2001**, *6*, 947 – 953.

[12] The EPSRC website is: www.epsrc.ac.uk.

[13] We have encouraged the PhD students and postdoctoral research assistants to apply their new methodology to the preparation of arrays with only partial success. However, when the new methodology has been applied to the preparation of arrays, it has proved a valuable learning experience for the students involved for many of the reasons outlined herein (especially in terms of working with more-polar molecules with more functional groups). The students begin to appreciate the multidisciplinary approach to drug discovery and that working in a team saves much time and effort in comparison to working by themselves. The experience of working in a team, the use of more automated equipment, and their introduction to the concept that in early lead optimization the synthesis of 2 mg of a compound for testing is more important than the overall yield provide the students with an industrial perspective. They also form valuable links with industrial colleagues; these connections should be useful to them in their future careers.

[14] For examples of the studies that have resulted from this initiative, see: "Palladium catalyzed tandem alkenyl- and aryl-C–N bond formation: a cascade N-annulation route to 4-, 5-, 6- and 7-chloroindoles", L. C. Henderson, M. J. Lindon, M. C. Willis, *Tetrahedron*, in press; "A direct route to triazole boronic esters and their application in the synthesis of small molecule arrays", J. Huang, S. J. F. Macdonald, A. W. J. Cooper, G. Fisher, J. P. A. Harrity, *Tetrahedron Lett.* **2009**, *50*, 5539 – 5541; "The rhodium carbene route to oxazoles: a remarkable catalyst effect", B. Shi, A. J. Blake, I. B. Campbell, B. D. Judkins, C. J. Moody, *Chem. Commun.* **2009**, 3291 – 3293; "Convergent synthesis of dihydroquinolones from *o*-aminoarylboronates", J. Horn, H. Y. Li, S. P. Marsden, A. Nelson, R. J. Shearer, A. J. Campbell, D. House, G. G. Weingarten, *Tetrahedron* **2009**, *65*, 9002 – 9007; "Convergent, Regiospecific Synthesis of Quinolines from *o*-Aminophenylboronates", J. Horn, S. P. Marsden, A. Nelson, D. House, G. G. Weingarten, *Org. Lett.* **2008**, *10*, 4117 – 4120; "Analysis of Neighborhood Behavior in Lead Optimization and Array Design", G. Papadatos, A. W. J. Cooper, V. Kadirkamanathan, S. J. F. Macdonald, I. M. McLay, S. D. Pickett, J. M. Pritchard, P. Willett, V. J. Gillet, *J. Chem. Inf. Model.* **2009**, *49*, 195 – 208; "Rearrangement Strategy for the Synthesis of 2-Aminoanilines", A. Porzelle, M. D. Woodrow, N. C. O. Tomkinson, *Org. Lett.* **2010**, *12*, 1492 – 1495; "Synthesis of Benzoxazolones from Nitroarenes or Aryl Halides", A. Porzelle, M. D. Woodrow, N. C. O. Tomkinson, *Org. Lett.* **2010**, *12*, 812 – 815; "Facile Procedure for the Synthesis of *N*-Aryl-*N*-hydroxy Carbamates", A. Porzelle, M. D. Woodrow, N. C. O. Tomkinson, *Synlett* **2009**, 798 – 802.

[15] The first report of new methodology carries considerable kudos and usually appears in high-impact journals. Frequently, however, it is sufficient to demonstrate that the methodology "works" in high yields on a limited range of substrates. Functionality which may interfere with the methodology can be omitted because the novelty factor of the methodology alone is generally sufficient for acceptance by referees. The subsequent demonstration of a refined protocol compatible with a wide range of functionality appears to be valued less, perhaps because it is less novel and less amenable to publication in high-impact journals. The pressure to produce high-impact publications may reduce the incentive for such studies to be carried out. Nonetheless, it is often a subsequent study which indicates the value of the new methodology to the medicinal chemist.

[16] The links to the NIH website describing this initiative and the centers are: www.nigms.nih.gov/Initiatives/CMLD; www.nigms.nih.gov/Initiatives/CMLD/Centers/.

[17] Frequently, the synthesis of many monomers and substrates is outsourced.

[18] Over this time period, the synthesis of very few arrays was outsourced. We have found it much quicker and logistically much easier for arrays to be made in-house.

[19] The medicinal-chemistry team charged with carrying out the lead-optimization process, which results in a clinical candidate, typically comprises (in decreasing order of seniority): a team leader (an experienced medicinal chemist and manager, usually with a PhD), two or three experienced medicinal chemists (graduates with or without a PhD), one or two less experienced graduates, and one or two university students on a one-year placement. All except the team leader carry out synthetic work;

all except the students contribute towards the medicinal chemistry commensurate with their grade and experience.
[20] S. M. Free, J. W. Wilson, *J. Med. Chem.* **1964**, *7*, 395 – 399.
[21] K. McClure, M. Hack, L. Huang, C. Sehon, M. Morton, L. Li, T. D. Barrett, N. Shankley, J. G. Breitenbucher, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 72 – 76.
[22] W. Zhai, N. Flynn, D. A. Longhi, J. A. Tino, B. J. Murphy, D. Slusarchyk, D. A. Gordon, A. Pendri, S. Shi, R. Stoffel, B. Ma, M. J. Sofia, S. W. Gerritz, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5083 – 5086.
[23] Historically, the majority of reactions were carried out on a less than 10 g scale owing to difficulties in purification; as a result, intermediate availability may have been limited. However (outside of cost constraints or limited commercial availability of the starting materials), intermediate availability is now less of an issue for two reasons: first, the advent of multiple vendors offering the cheap synthesis of bespoke intermediates, and second, the ready availability of purification systems which can readily handle 50–100 g of crude material.