ELSEVIER

feature

The genetics of a pharma merger

Philippe Sanseau[1], Marie Chabot-Fletcher[2] and Michael J. Browne[3*], mjbrowne@btconnect.com

The 1990s and early years of this century have seen a series of large-scale mergers and acquisitions in the Pharmaceutical and Biotech arena. These activities each required integration at multiple levels. One of the most important activities is the integration of the R&D pipelines of the participants. We outline the combined portfolio and bioinformatic strategy that was used, and detail the lessons learned for the longer term, from the GlaxoWellcome–SmithKline-Beecham merger in 2000. To date, this has been the largest merger of two equally sized Pharma R&D organisations.

Increased opportunities pose organisational challenges for strategic mergers

The combined effect of the molecular biology revolution of the 1970s and 1980s, and the genomics revolution of the 1990s, highlighted with the completion of the human genome sequence [1], has given biomedical research access to an increasing number of genes encoding potential 'druggable' targets. A consensus number of targets for the entire compendium of currently approved drugs has been estimated at only 324 [2], while it is proposed that there may be 2000–3000 potential druggable targets in the human genome [3]. The increased potential for new drug discovery is important although there remain concerns that too few of the targets from the genomics era have been clinically validated and it is clear that overall pipeline attrition remains a key issue for the industry [4].

At the time of the GlaxoWellcome (GW) and SmithKline-Beecham (SB) merger we needed to document and merge two large R&D portfolios into a single new entity that was organised and focused to meet the goals of the newly formed company GlaxoSmithKline (GSK). Given the concerns about the apparent shortage of validated targets and considering that the human genome and expressed sequence tags (ESTs) [5] had been entering the public domain incrementally from the mid-1990s onwards, we hypothesised that both companies were likely to have initiated work on an overlapping set of targets. Hence, there would be redundancy between the two premerger portfolios of human targets – particularly in the early research phases. We anticipated that duplicated projects could be identified, eliminated or integrated to accelerate progress by virtue of combining datasets and knowledge. We knew early on that each R&D portfolio consisted of several hundred research programmes as well as (the publicly reported) development projects in many therapeutic areas located in over a dozen R&D sites worldwide. This would probably require a considerable effort to annotate, analyse, prioritise and integrate.

Using a scientific approach to organise science

It was reasoned that differing scientific, corporate and regional nomenclatures and staff movements might compound problems of comparison and integration exactly when time was a major constraint on decision making for the new company. Rather than implement a new system for nomenclature we took a different approach. We employed the universal language of genetics (DNA sequences) to annotate and organise the reassembly of the premerger portfolios into an integrated portfolio. This relied on uniquely tagging all projects, where possible, with the appropriate public domain gene identification number (ID) from the gene sequence for the associated human pharmacological targets in each of the premerger R&D portfolios. Of particular concern was the need to avoid potentially expensive 'false negatives' arising from inconsistent nomenclature. That issue could be illustrated simply with well-known drug targets such as phosphodiesterase 4A (PD4A) or a potassium channel such as KCNK1. In the public domain alone, other known aliases for PD4A are PDE46, PDE4A4B, PHOSA and DUNCE. In the case of KCNK1 at least seven additional names are known: TWIK-1, TWIK1, HOHO1, KCNO1, HOHO, K2p1.1 and DPK. We also realised that the same pharmacological target might be

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2008.12.001

managed independently in two or more different therapeutic areas at the same time for different indications.

We associated the gene IDs from the portfolio projects of the two separate companies with their corresponding gene sequence. This was done partly computationally and partly manually. The sequences used in our analysis represented targets that had moved beyond 'omic' scale investigations (bioinformatics, microarray and proteomics) used for early target identification work into specific projects of work. A bonus side benefit was that the relatively small number of cell-based or phenotypic assay-based programmes where no specific molecular target had been identified was simple to identify and manage manually.

The approach that we adopted was, we think, unorthodox. This may have been influenced by the unprecedented scale of the task and the organisational structure. At the time of the project all three authors were working as members of a flexible, *ad hoc*, multidisciplinary portfolio merger team. Although each author belonged nominally either to a portfolio or a bioinformatics department in either company, all authors were familiar with gene sequence analysis as a result of earlier career paths, thus aiding rapid coordination.

Comparing and merging the new portfolio

The gene sequences were then used in a rapid, simple Blast search [6] to identify portfolio overlaps (Fig. 1). The targets and associated gene IDs of the GW and SB portfolios were separately preallocated to defined, agreed R&D phases (Fig. 2 – footnote contains phase definitions for the early portfolio). The portfolio 'fusion' was accomplished rapidly and with a high level of confidence and transparency. In the combined early portfolio we identified a total of 328 unique targets. These targets were divided into three phases: target to hit (T2H); hit to lead (H2L) and lead optimisation (LO). In the T2H phase there were a total of 202 unique targets and of these 17 represented premerger duplications. Interestingly only six of these focused on the same indication. Largely as we had expected, 11 others focused on alternative indications in areas that were as divergent as neurology versus chronic obstructive pulmonary disease and restenosis versus cancer (not shown). In the H2L phase there were a total of 55 unique targets but there were no overlaps within that phase. In the LO phase there were only three overlaps out of a total of 71 targets. In addition to the phase-specific duplications, there were a few target overlaps where one company had advanced their project by a phase or two beyond the other. In summary, the strategy that we had employed easily identified 37 areas of target overlap among hundreds of projects distributed worldwide – as had been intended.

From another perspective, we were surprised that there were comparatively few potential overlaps between the two portfolios. Thus, despite the increasingly level playing field of publicly available genomic-scale information (especially, human genome and ESTs) from the mid-1990s onwards, the operational 'genomes' of each premerged company had evolved to be quite different. It seems apparent that, even with the same basic public domain information on potential targets in the human genome, the choices about where to put priority efforts that are made by different scientific and management teams are quite diverse. Each, presumably, was strongly influenced by the understanding of biology (target validation), the available chemical (including *in silico* or biopharmaceutical) space, expertise, perception of patient needs and recruitment strategies for staff and scientific management.

Long-term impact and value

What has been the long-term impact of this novel, large-scale exercise? It is difficult to assess directly the impact of merging the two portfolios over multi-year R&D timescales because so many strategic and scientific factors (e.g. attrition and turnover) are involved; however, it is noteworthy that the number of new chemical entities (NCEs) in development in Phase 2 increased from 15 in 2001 to 45 in 2004, and by 2008 GSK had the largest late-stage development portfolio in the industry [7].

Aside from the merger-related utility in this instance, our later day-to-day experience also suggests that having an R&D portfolio that is annotated with agreed standard gene IDs and sequences increases confidence in the data and simplifies communication and analysis in an

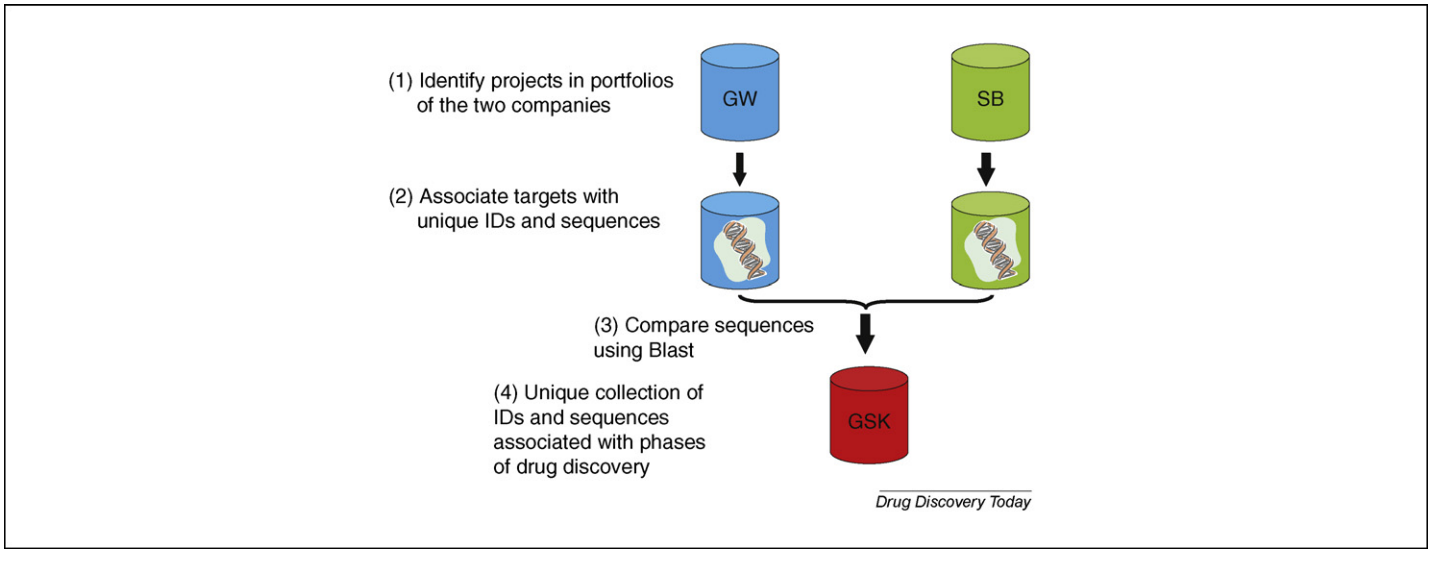

FIGURE 1

Assignment and integration of targets and associated projects in the merging R&D portfolio. (1) and (2) Unique sequences were associated for both portfolios by using, when available, public domain IDs (e.g. SwissProt) associated with targets or by talking to appropriate scientists when specific GW or SB IDs were only available. (3) Sequences were compared using the Blast search tool. (4) Unique sequences were selected, mapped to phases of drug discovery and integrated into a single GSK resource for decision making.

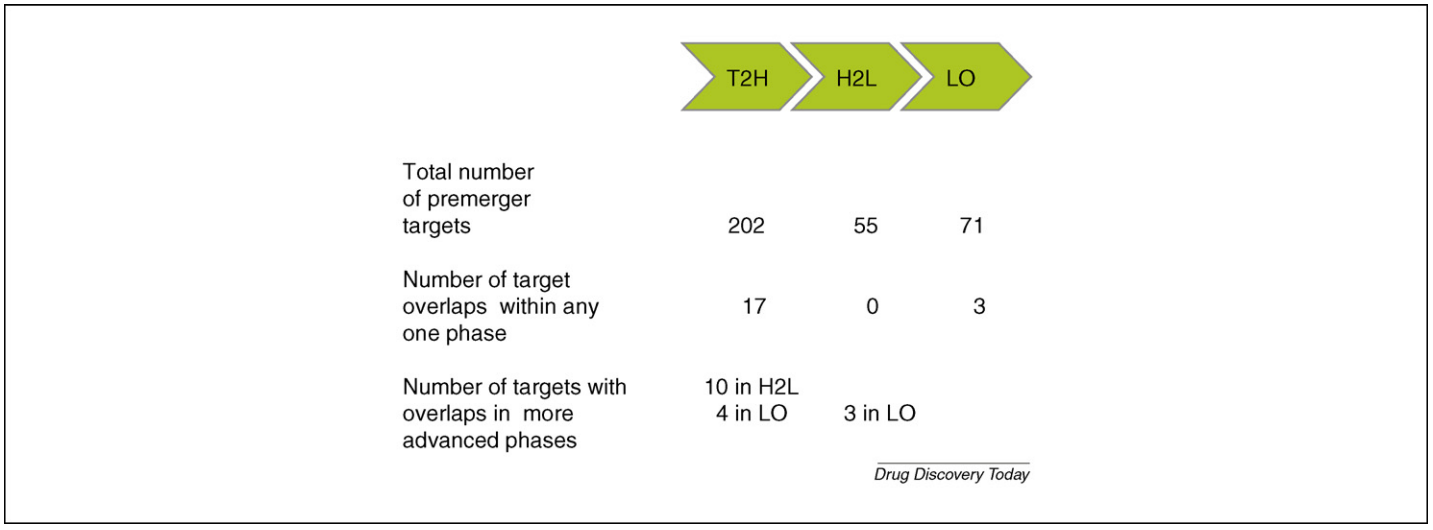

FIGURE 2

Overlap of targets in each phase of the combined GSK portfolio. T2H: target to hit – phase starts at selection of target for protein expression, assay construction and screening (including high throughput), computational and cheminformatic analysis up to identification of first round of hits. H2L: hit to lead – triage of hits, array-based expansion of initial hit series, initial structure activity exploration, evaluation of potential for development and selection of one or more series. LO: lead optimisation – more detailed structure activity analysis of lead series (e.g. computational chemistry, potency and selectivity profiling of modified structures, p450 activity, *in vivo* pharmacokinetics and pharmacodynamics) before development candidate selection.

increasingly complex R&D environment. This should be of benefit in a conventional, large, centrally managed enterprise, as well as the localised model of Centres of Excellence for Drug Discovery (CEDD) adopted by GSK and in more recent models emphasising the need for a network of collaborations and alliances with external partners. It is worthwhile noting that Pharma companies are becoming more comfortable with using standards defined by the public domain: for example HUman Genome Organisation (HUGO) standards for gene names [8].

Concluding remarks

We have outlined a combination portfolio and bioinformatic strategy that was used successfully to facilitate the merger of two large pharmaceutical R&D organisations. We propose that the benefit of a simple and standardised gene-based annotation of projects holds true both at a time of major reorganisation (e.g. a merger) or when considering collaborations or alliances. The benefits of this simple combination approach will probably grow as the number of pharmaceutically druggable targets is itself increasing. Alliances and collaborations have been a vital feature of pharmaceutical, biotech and academic innovation for some time, and seems likely to continue and even diversify.

Acknowledgement

The authors would like to thank the members of the GlaxoSmithKline Portfolio merger teams.

References

1 The White House, Office of the Press Secretary (2000) *President Clinton announces the completion of the first survey of the entire human genome*, June 25th 2000
2 Overington, J.P. *et al.* (2006) How many drug targets are there? *Nat. Rev. Drug Discov.* 5, 993–996
3 Russ, A.P. and Lampel, S. (2005) The druggable genome: an update. *Drug Discov. Today* 10, 1607–1610
4 Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3, 711–715
5 Lee, Y. *et al.* (2005) The TIGR gene indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Res.* 33, D71–D74
6 Altschul, S.F. *et al.* (1990) Basic alignment search tool. *J. Mol. Biol.* 215, 403–410
7 Garnier, J.P. (2008) Rebuilding the R&D Engine in Big Pharma. *Harvard Business Rev.* 86, 68–76
8 Bruford, E.A. *et al.* (2008) The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res.* 36, D445–D448

Philippe Sanseau
Molecular Discovery Research,
GlaxoSmithKline, Stevenage SG1 2NY,
United Kingdom

Marie Chabot-Fletcher
Birchrunville, PA 19421, USA
mchabot@ix.netcom.com

Michael J. Browne
M J Browne Consultancy Ltd,
PO Box 440, Welwyn Garden City,
Hertfordshire AL7 9EE,
United Kingdom
mjbrowne@btconnect.com