Analysis of the reactions used for the preparation of drug candidate molecules

John S. Carey,*[a] David Laffan,[b] Colin Thomson[c] and Mike T. Williams[d]

Received 16th February 2006
First published as an Advance Article on the web 3rd May 2006
DOI: 10.1039/b602413k

The purpose of this perspective is to indicate the range of chemistries used in the manufacture of drug candidate molecules and to highlight certain gaps in current technologies. To do this a survey was carried out of chemical syntheses within the Process Chemistry R&D departments of GlaxoSmithKline, AstraZeneca and Pfizer.

[a]*GlaxoSmithKline Pharmaceuticals, Old Powder Mills, nr. Leigh, Tonbridge, UK. E-mail: john.s.carey@gsk.com; Fax: 44 (0)1732 372355; Tel: 44 (0)1732 372245*
[b]*AstraZeneca PR&D, Silk Road Business Park, Macclesfield, Cheshire, UK SK10 2NA. E-mail: dave.laffan@astrazeneca.com; Fax: 44 (0)1625 500780; Tel: 44 (0)1625 517483*
[c]*AstraZeneca R&D Charnwood, Bakewell Road, Loughborough, Leics, UK LE11 5RH. E-mail: colin.thomson@astrazeneca.com; Fax: 44 (0)1509 645588; Tel: 44(0)1509 644273*
[d]*Pfizer Global R&D, Sandwich, Kent, UK CT13 9NJ. E-mail: mike.t.williams@pfizer.com; Fax: 44 (0)1304 655586; Tel: 44 (0)1304 646383*

Introduction

The Process Chemistry R&D departments of pharmaceutical companies have the following roles: the supply of drug substances for development as well as the discovery and scale-up of syntheses suitable for commercial manufacture. The scale of manufacture ranges from kilograms to tonnes. The targets of these syntheses are pre-set by the candidates selected for development from Medicinal Chemistry. These are generally small molecules and are rarely derived from natural products (<5% of the survey).

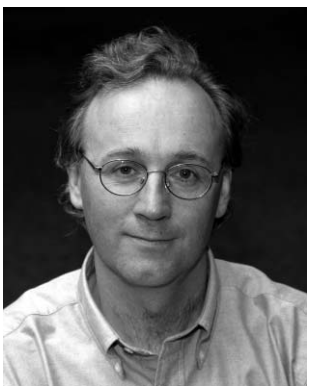
John Carey received a BSc in Chemistry from Edinburgh University. He moved to University of Manchester to study a PhD under the supervision of Professor E. J. Thomas on remote asymmetric induction using allylstannanes. He did postdoctoral research with Professor G. A. Molander at the University of Colorado on the application of [4 + 3] annulation reactions to natural product synthesis. He worked for Ciba Central Research/Novartis in Macclesfield for two years prior to joining the Synthetic Chemistry Department of SmithKline Beecham (now GlaxoSmithKline) at Tonbridge in 1997.

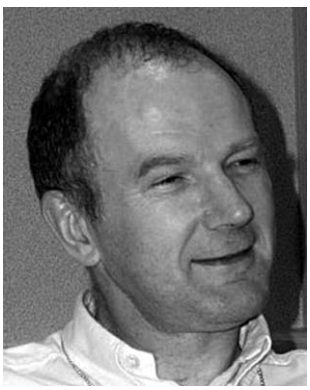
David Laffan obtained his BA from Cambridge University, where he stayed to complete a PhD under the supervision of Ian Paterson on the synthesis of the macrocyclic skeleton of erythromycin. He started work in the fine chemicals industry with Lonza in Switzerland immediately after his PhD. After six years he returned to the UK and moved to the pharmaceuticals industry by joining Zeneca (now AstraZeneca) in the Process R&D department in 1995.

Colin Thomson graduated from Glasgow University with 1st Class Honours in Chemistry in 1968. He then undertook post-graduate research at Glasgow, obtaining his PhD in 1971. His first appointment was as a medicinal chemist at Pfizer Central Research in Sandwich, from 1971–1978, spending 1974 at Groton in the USA. He then moved to Process R&D at Reckitt and Colman until 1981 when he moved to Fisons Pharmaceuticals in Loughborough. He remained there through the Astra takeover and AstraZeneca merger. He is currently a Principal Scientist in Process R&D.

Mike Williams graduated from King's College London with a BSc in Chemistry. Following a year spent working in Medicinal Chemistry for ICI Pharmaceuticals (subsequently Zeneca) he completed a PhD under Professor Charles Rees at Liverpool University on reactive intermediates and heterocyclic synthesis. In 1972 he joined the Chemical Research and Development department of Pfizer at Sandwich, a group that he now heads.


John S. Carey


David Laffan


Colin Thomson


Mike T. Williams

There have been recent reviews written by Process R&D chemists describing the rationale for route selection[1] and a historical survey of reactions carried out in a single pilot plant.[2] This perspective complements these reviews by providing a pan-industry cross site analysis of the chemistry currently used in drug development, informing researchers about the needs of the pharmaceutical industry and indicating significant gaps in synthetic methodologies.

Methodology and data set

In carrying out this survey, the syntheses of 128 drug candidate molecules were analysed, these were divided between the three companies and covered all therapeutic and geographic areas that the companies have R&D interests in, Table 1.

The synthesis of each molecule was analysed and the reactions placed in the following reaction categories; acylation, aromatic heterocycle formation, C–C bond formation, deprotection, functional group addition (FGA), functional group interconversion (FGI), heteroatom alkylation & arylation, oxidation, protection, reduction, resolution and miscellaneous, Table 2. Salt formation/breaking and recrystallisations have not been included in the analysis. These categories reflect the chemical transformation and not the technology/equipment requirement. Some transformations could be classified in more than one category. Each transformation has only been entered once into the data set, but may be discussed in more than one section.

This perspective is a snapshot of compounds in development, which inevitably means that it is heavily biased towards compounds in early development (*ca.* 50% phase I or earlier, requirements of drug candidate <10 Kg). The early phase of development tends to use syntheses that are modifications of discovery routes, which includes many non-constructive steps (protection/deprotection), less reliance on crystallisation for purification and less interest in environmental and economic concerns. This data set reflects the chemistry employed during development and not the chemistry employed in full production.

Despite the wide spread in therapeutic and geographic areas between and within the companies, there is a surprising degree of similarity in the patterns emerging from the data. The data are therefore presented as a combination from the three companies as there is little to be gained from looking at each company individually.

Overview of pharmaceutical synthesis

The drug substances discussed here are all small molecules (<550 MW), >90% of which contain a nitrogen atom, >90% contain an aromatic ring and 54% are chiral. This shows the nature of the compounds coming from medicinal chemistry programmes into the Process Chemistry R&D departments of the three companies. The preparation of biopharmaceuticals falls outside the scope of this review.

In considering drug substance syntheses there are several constraints. First among these is the requirement for reproducible high quality product; certain impurities need to be controlled to very low levels (*e.g.* heavy metals <10 ppm, genotoxic agents). This does not preclude the use of any methodology, provided suitable attention is given to its position in the synthesis and control measures are in place. Simple, clean synthetic manipulations as

Table 1 Headline data

	AstraZeneca	GlaxoSmithKline	Pfizer	Total
Number of syntheses	45	39	44	128
Total number of chemical transformations	371	310	358	1039
Average number of chemical transformations per synthesis	8.2	7.9	8.1	8.1
Number of chiral compounds	25	23	21	69
Number of chiral centres	46	52	37	135
Number of chiral centres generated	22	19	20	61
Number of substituted aromatic starting materials	64	79	63	206
New aromatic heterocycles formed	14	11	29	54

Table 2 Summary of reaction categories

Reaction category	AstraZeneca	GlaxoSmithKline	Pfizer	Total	Percentage of total reactions
Heteroatom alkylation & arylation[a]	87	57	52	196	19%
Acylation[a]	41	37	50	128	12%
C–C bond forming[a]	31	41	44	116	11%
Aromatic heterocycle formation[a]	16	10	26	52	5%
Deprotection[b]	54	56	49	159	15%
Protection[b]	18	16	27	61	6%
Reduction[b]	27	24	43	94	9%
Oxidation[b]	17	7	16	40	4%
Functional group interconversion[b]	43	34	27	104	10%
Functional group addition[b]	13	8	12	33	3%
Resolution[b]	14	8	8	30	3%
Miscellaneous	10	12	4	26	3%
Totals	371	310	358	1039	

[a] Reactions used for molecular construction. [b] Modifying reactions.

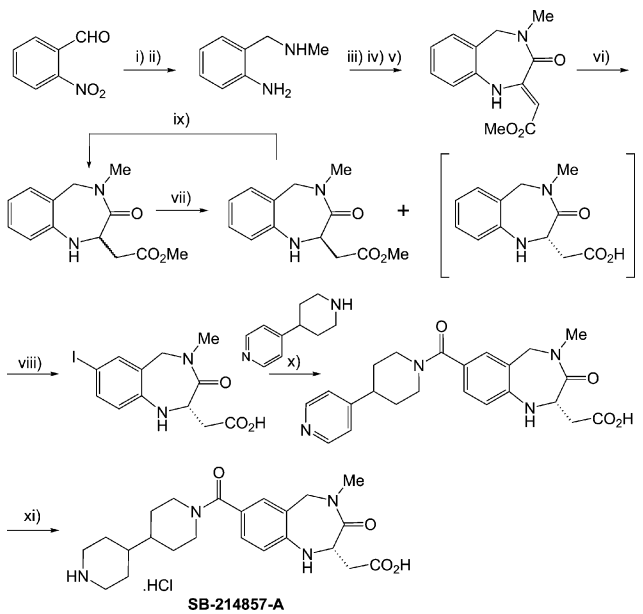
the final steps in the synthesis are highly desirable. Regulatory constraints, both environmental and pharmaceutical, influence the type of chemistry selected. Processes also have to be economic to operate. Pharmaceutical companies tend to have multi-purpose batch reactors, which leads to a preference for processes that can be accommodated in this type of equipment.[1]

These constraints lead to a preference for modular convergent syntheses, *i.e.* simple construction from complex fragments. For example, it is common practice to purchase chiral synthons rather than introduce chirality late in the synthesis.

For the supply of these complex fragments and in dealing with some of the technology constraints, the pharmaceutical industry works in close partnership with the fine chemicals industry.

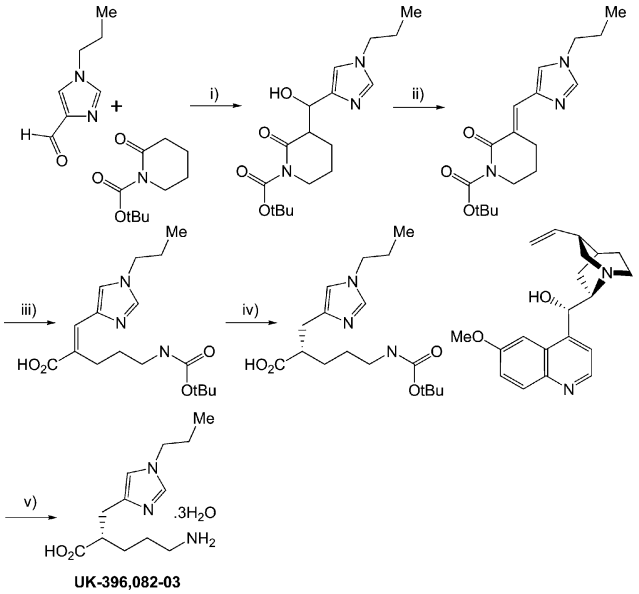
Typical examples of process chemistry R&D syntheses

Scheme 1 shows the synthesis of **SB-214857-A**, a potential drug candidate molecule, the development of which was halted during phase III clinical trials, but by that stage the route of manufacture of the drug substance had been established and was totally different from the discovery route.[3] The manufacturing route shown contains 11 chemical transformations of which eight are promoted by catalysts. There are no classical protections or deprotections, the pyridine group acts as a masked piperidine. There are three reductions and no oxidations. The chirality is generated by an enzyme catalysed dynamic resolution and the unwanted enantiomer is recycled. The overall yield is 25%.



Scheme 1 *Reagents and conditions*: i) $MeNH_2$, $NaBH_4$, MeOH, H_2O. ii) H_2, Pd–C, MeOH 97%. iii) Dimethyl acetylene dicarboxylate, MeOH. iv) NaOMe, MeOH. v) AcOH, MeOH, 68%. vi) NH_4CO_2H, Pd–C, MeOH, 93%. vii) Novozym 435, *t*-BuOH, H_2O, NH_3, pH 7.0. viii) Pyridine iodine monochloride complex, H_2O, NaOH, pH 7.0, 38%. ix) NaOMe, MeOH, $(MeO)_2CO$, 38% (over two steps). x) $PdCl_2(PPh_3)_2$, CO, 4,4′-pyridylpiperidine, anisole, 86%. xi) H_2, Pd–C, iPrOH, then $C_5H_5N{\cdot}HCl$, EtOH, CH_2Cl_2, H_2O, 78%.

Scheme 2 depicts the 5 step synthesis of **UK-396,082-03**, a relatively small and highly polar compound which is in early development as an anti-thrombotic agent.[4] This development synthesis differs from the discovery synthesis in two key respects. Firstly the route was modified and reordered to introduce crystalline process intermediates to aid isolation and purification. Secondly the original use of chiral HPLC to resolve the penultimate intermediate was replaced by an asymmetric hydrogenation to introduce the chiral centre, with crystallisation of the quinidine salt purging residual levels of the unwanted enantiomer.



Scheme 2 *Reagents and conditions*: i) LHMDS, THF, −50 °C, 82%. ii) Et_3N, MsCl, CH_2Cl_2, 91%. iii) NaOH, H_2O, THF, then AcOH, iPrOH, 75%. iv) Quinidine, MeOH, 150 psi H_2, [(*R*,*R*)-iPr-5-FcRhCOD]BF_4, then EtOAc, 75%. v) NaOH, then HCl then Dowex column then acetone, 76%.

These two syntheses highlight some important features of modern process chemistry; use of catalysts to promote both asymmetric and non-asymmetric reactions, minimal use of protecting groups and the acceptability of isolated enzymes.

Headline data

Number of synthetic steps

The 128 syntheses contained 1039 chemical transformations, an average of 8.1 steps (GSK 7.9, AZ 8.2, Pfizer 8.1). The distribution of stages per compound is shown in Fig. 1.

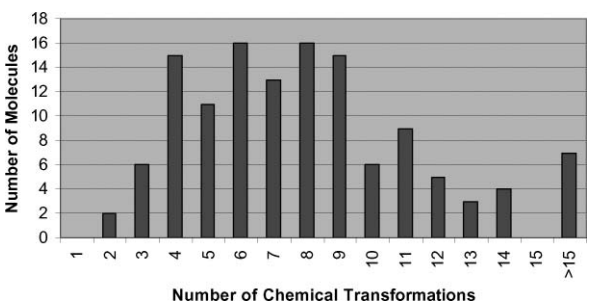


Fig. 1 Number of chemical transformations per compound.

Chirality

Of the 128 molecules analysed, 69 (54%) are molecules containing at least one stereogenic centre, Fig. 2. Of the 69 chiral molecules 67 are being developed as single stereoisomers, with only two as racemates.

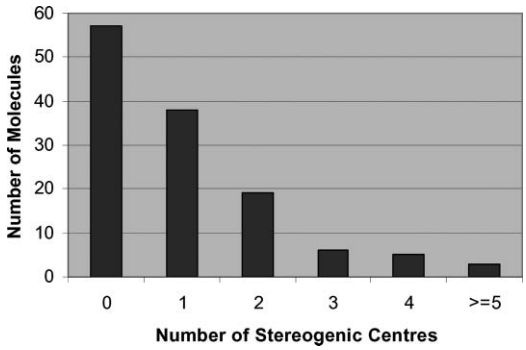


Fig. 2 Number of stereogenic centres per molecule.

Aromatics

There are 206 substituted aromatic or heteroaromatic starting materials used for the 128 compounds and 54 new heteroaromatics are constructed.

Construction *vs.* modification

Chemical transformation can be broadly classified into two types, constructive and modifying. The constructive categories are: acylations, aromatic heterocycle formation, C–C bond formation, heteroatom alkylation & arylation and some of the miscellaneous. As can be seen from Table 2 the breakdown for reactions that are involved in molecular construction is approximately 48%. The modifying transformations are: deprotection, functional group addition (FGA), functional group interconversion (FGI), oxidation, protection, reduction, resolution and some miscellaneous.

Chirality

The requirement to produce chiral compounds as single enantiomers/diastereoisomers for pharmaceutical applications is well understood and a recent publication has provided a snapshot of asymmetric methods used on a process scale.[5] A typical target value for the enantiomeric purity is 99.5%.

The source of the 135 chiral centres in 69 chiral drug syntheses was surveyed and is shown in Fig. 3. For 27 of the chiral molecules all the chirality resided in the starting materials purchased. For 30 molecules all the chirality was generated during the syntheses (43 stereogenic centres generated) and for a further 12 molecules there was a mix of purchased and generated chirality (a further 18 stereogenic centres generated).

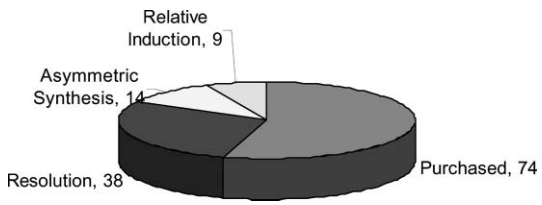


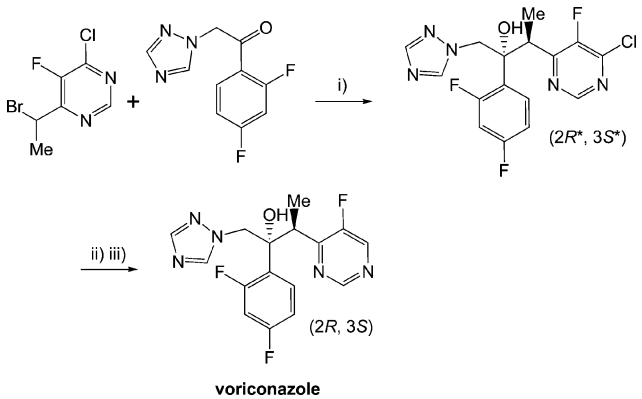
Fig. 3 Source of chiral centres

It is evident from the data that the predominant strategy is to purchase chiral starting materials/intermediates from the fine chemical industry.[6] This is driven by the desire to introduce chirality early in a synthesis, consistent with the assembly of smaller complex fragments and determined by their availability.

The methodologies used to generate the 74 chiral centres purchased are not covered in the survey and the breakdown may well be different to the 61 chiral centres generated in the later stages of the syntheses.

The predominant method of in house generation of chirality is resolution, with two thirds of these being performed by classical salt formation and the remainder being evenly distributed between dynamic kinetic, chromatographic and enzymatic methods. The availability of screening methods to develop classical resolutions, the increased understanding of crystallisations and the ease of scale-up continue to make this the preferred methodology for many chiral molecules.[7]

When a chiral centre is generated later in the synthesis within a more complex substrate it seems that few methods exist that are sufficiently straightforward to be operated economically. The preferred methods are relative diastereocontrol or preparation of a racemate followed by resolution. This is exemplified by the synthesis of voriconazole, Scheme 3.[8] A diastereoselective Reformatsky reaction was employed to establish the relative stereochemistry followed by camphorsulfonic acid resolution to obtain enantiomerically pure voriconazole.



Scheme 3 *Reagents and conditions*: i) Zn, I_2, THF, 90%. ii) H_2, Pd–C, NaOAc, EtOH, 85%. iii) (1*R*)-10-Camphorsulfonic acid, MeOH then *aq* NaOH, 40%.

Asymmetric synthesis only accounts for a smaller proportion, approx. 20%, of the chiral centres generated, but it is noteworthy that the methods applied are catalytic in nature. Even for a well developed methodology, such as catalytic asymmetric hydrogenation, application to a moderately complex substrate rarely yields the target enantiomeric purity directly. The synthesis of **UK-396,082-03** shown in Scheme 2 is a good example of the use of asymmetric hydrogenation.

Substituted aromatic starting materials

A wide variety of di-, tri- and tetrasubstituted benzenoid aromatics are purchased as starting materials as outlined in Table 3.

Table 3 Polysubstituted benzenoid starting materials

Substitution pattern	Number of examples	Frequency
1,2-Ph	16	11%
1,3-Ph	17	12%
1,4-Ph	38	27%
1,2,3-Ph	13	9%
1,2,4-Ph	46	32%
1,3,5-Ph	5	4%
1,2,3,4-Ph	4	3%
1,2,3,5-Ph	2	1%
1,2,4,5-Ph	1	1%
Total	142	100%

While the most common of these have been prepared using standard electrophilic substitution reactions, a number have synthetically challenging substitution patterns and are hence very expensive. Fig. 4 shows examples of tri- and tetrasubstituted aromatics These are seemingly simple starting materials that are often very difficult to produce. New methods for the synthesis of these difficult substitution patterns would be welcomed.

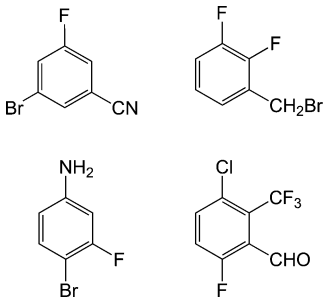


Fig. 4 Difficult to access substituted aromatic starting materials.

Heterocycle occurrence and formation

From the data set of 128 compounds, preformed aromatic heterocyclic starting materials were purchased in 59 cases, while 54 aromatic heterocycles were synthesised. The formation of saturated heterocycles is captured under the relevant bond-forming step (*e.g.* acylation or alkylation). There is a strong preponderance of N-containing heterocycles, as shown in Table 4. (Note that some heterocycles contained an O or an S in addition to N, *e.g.* oxazole, so that the frequency exceeds 100%.)

Within the purchased heterocyclic starting materials 43 were electron-deficient 6-membered heterocycles, 1 was a fused 5/6 bis-heterocyclic system, and 15 were electron-rich 5-membered heterocycles. The 6 most commonly occurring heterocyclic systems in this survey are shown in Table 5.

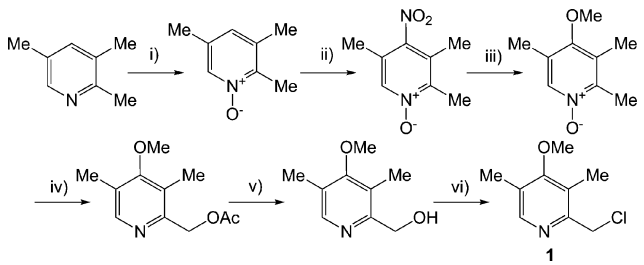
Table 4 Heterocycles purchased or synthesised, classified by heteroatom

Purchased heterocycles	Number of examples	Frequency
N-Containing	54	92%
O-Containing	4	7%
S-Containing	3	5%
Total	63	
Synthesised heterocycles	**Number of examples**	**Frequency**
N-Containing	53	98%
O-Containing	10	19%
S-Containing	5	9%
Total	68	

Table 5 Most commonly occurring aromatic heterocycles

Heterocycle	Number purchased	Number synthesised	Total
Pyridine	23	3	26
Quinazoline	12	5	17
Pyrazole	3	5	8
Pyrimidine	4	3	7
1,2,4-Triazole	0	7	7
Thiazole	1	4	5

Of the purchased pyridines 6 were mono-, 11 were di-, 5 were tri- and 1 was tetrasubstituted, with 2,5-disubstitution the most common pattern (9 examples). The maturity of the market for the supply of substituted pyridines was such that in only 3 of these cases was it necessary to introduce a further ring substituent to attain the substitution pattern required in the target. Even though pyridine chemistry is well understood, some substitution patterns can still be difficult to prepare, *e.g.* the pyridine **1**, a starting material for omeprazole, Scheme 4.[9]



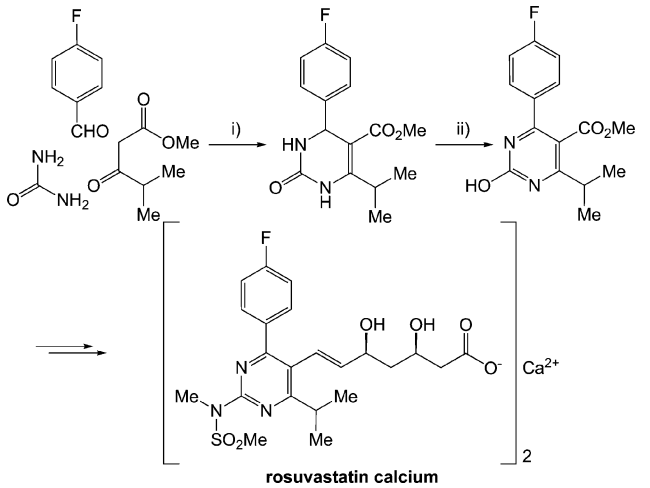
Scheme 4 *Reagents and conditions*: i) H_2O_2, Ac_2O. ii) HNO_3, H_2SO_4. iii) NaOH, MeOH. iv) Ac_2O, AcOH. v) NaOH, H_2O. vi) $SOCl_2$.

The synthesised heterocycles comprised 36 monocycles, 17 bicyclic compounds and 1 tricyclic compound. Of the 54 ring-forming steps, 34 prepared electron-rich 5-membered rings, 17 formed electron-deficient 6-membered rings, and 3 made 7-membered 1,4-diazepines. Thus there was a pattern where about 70% of the 5-membered rings were synthesised, and a very similar proportion of 6-membered rings were purchased. Mechanistically, most of the ring-forming reactions were classified as either condensations (24 examples) or cyclodehydrations (12 examples). In only 4 instances were the heterocycles formed by a cycloaddition reaction. Scheme 5, exemplifying the formation of a pyrimidine, shows the general trend in the use of established condensation chemistry.[10]

Protections and deprotections

Protections and deprotections account for 6% (61 reactions) and 15% (159 reactions), respectively, of the total chemical transformations and therefore between them account for >20% of all the transformations and almost 2 chemical transformations per molecule. The reason that deprotections significantly outnumber protections is that many of the starting materials purchased contain the protecting groups in place.

On a positive note, 45 out of the 128 syntheses (35%) were achieved without the use of protecting groups. The average number of synthetic steps in the syntheses achieved without the use of



Scheme 5 *Reagents and conditions*: i) $CuCl_2$, H_2SO_4, MeOH, 85%. ii) HNO_3, 94%.

protecting groups was 5.9, which is considerably shorter than the overall average of 8.1. The longest synthesis achieved without protecting groups was 12 steps.

Protections fall predominately into three groups: protection of the amino NH group (39%), protection of a carboxylic acid as an ester (28%), and protection of a hydroxyl group (30%), Table 6. The most common protecting group for the amino NH group is BOC, followed by a combination of benzylamine/CBZ group, *i.e.* groups that can be removed by hydrogenation/hydrogenolysis. In general the carboxylic acid group is protected as a simple alkyl ester such as a methyl ester or an ethyl ester. A wider variety of protecting groups was used for the protection of hydroxyl groups. The most common groups were benzyl ether, silicon-containing protecting groups and acetate esters. Higher molecular weight protecting groups, such as trityl, are disfavoured as they are costly and reduce throughput.[1]

The distribution of functional groups that are deprotected closely mirrors those protected, Table 7. Hydrolysis of esters and amides to carboxylic acids (29%), deprotection of the amino NH group (47%), deprotection of the hydroxyl group (14%) and deprotection of thiol SH (3%) are the most common. The most common deprotections of the amino NH are BOC and the benzylamine/CBZ group. The most common deprotections for the hydroxyl group are benzyl ethers, deprotection of phenolic ethers to give phenols and the removal of silicon-containing protecting groups.

Table 6 Protections

Functional group	Number of examples	Frequency
Amino	24	39%
N–BOC	*12*	
N–Bn/Cbz	*5*	
Hydroxyl	18	30%
O–Bn	*3*	
O–SiR$_3$	*3*	
O–Ac	*2*	
Carboxylic acid	17	28%
Other	2	3%
Total	61	

Table 7 Deprotections

Functional group	Number	Frequency
Amino	74	47%
N–Boc	*32*	
N–Bn/Cbz	*21*	
Carboxylic Acid	46	29%
Hydroxyl	23	14%
O–Bn	*8*	
Ar–OR	*7*	
O–SiR3	*3*	
Thiol	3	2%
Other	13	8%
Total	159	

The use of silicon protecting groups may seem small, but they generally contribute to a lack of crystallinity in the intermediates and, except for TMS, they can be expensive.

Although there exist many different protecting groups[11] for the protection and deprotection of multivarious functional groups only a small number is routinely used in the preparation of active pharmaceutical ingredients. The choice of protecting group for the amino NH is dominated by two strategies, use of BOC which is removed under acidic conditions and use of benzylamine/CBZ, which are removed by hydrogenation/hydrogenolysis. The most common protecting group strategy for the hydroxyl group is the use of benzyl ethers, which are removed by hydrogenation. Carboxylic acids are routinely protected as simple aliphatic esters.

Even a straightforward removal of a BOC group from a molecule containing aromatics/nitrogens can result in unacceptable levels of impurities resulting from incorporation of the *t*-Bu cation. By-products from the deprotection must also be benign. For example, HBr deprotection of a ArOMe results in the formation of MeBr, which requires tight control.

By and large protections and deprotections are designed out of a synthesis, *cf.* the synthesis of **SB-214857-A**, Scheme 1, where the development route contained no protecting groups whereas the original discovery route contained 5 protection/deprotection steps.[3]

Acylation

Acylation reactions, especially of nitrogen, are frequently used in the preparation of drug candidate molecules; the categories of acylation, and the occurences of each, are summarised in Table 8. Acylations comprised 12% of the reactions in this survey, and the dominance of *N*-acylations is to be expected from the frequency with which amides occur in drug molecules. Of the 53 small

Table 8 Acylations

Acylation	Number of examples	Frequency
N-Acylation to amide	84	66%
N-Sulfonation to sulfonamide	12	9%
N-Acylation to urea	8	6%
Carbamate/carbonate formation	7	5%
Amidine formation	5	4%
O-Acylation to ester	5	4%
Other	7	5%
Total	128	

Table 9 *N*-Acylation methods

Method	Number of examples	Frequency
Acid chloride	37	44%
Coupling reagent	21	25%
Mixed anhydride	11	13%
Carbonyl diimidazole	9	11%
Other	6	7%
Total	84	

molecule drugs whose sales in 2003 exceeded $1 Bn, 9 contained an amide and a further 3 a sulfonamide moiety.[12]

Within the amide formation category, the breakdown of *N*-acylation methods is tabulated in Table 9. Acid chlorides were by far the most common acylating agent, in Table 9 all 37 examples involved *in situ* formation of the acid chloride, while in a further 14 cases the acid chloride was isolated and counted as a separate step (see FGI section). The use of mixed anhydrides provides an inexpensive and readily scaled process for *N*-acylation that is particularly valuable for cases prone to epimerisation. The moderately priced reagent carbonyl diimidazole (CDI) is gaining popularity, as reactions are readily scaled up and worked up.[13] Although amide formation with coupling reagents such as carbodiimides are frequently used for early development, for later development these reagents are usually developed out of processes as they are sensitisers and are relatively costly.

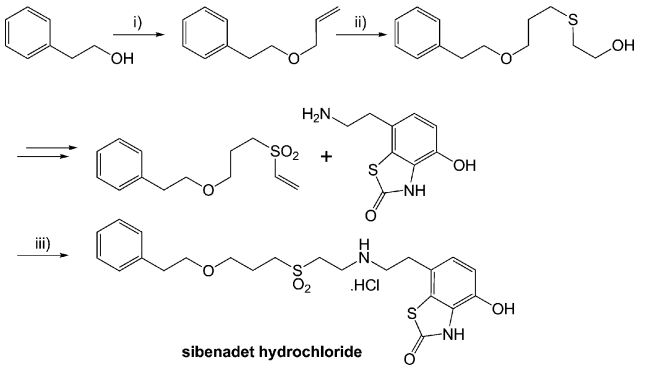
Although economic amide formation processes are available (notably the acid chloride and mixed carbonic anhydride methods), these are not particularly atom efficient or "green". Significantly, there was not a single example of a catalytic amide formation in the survey—although enzymatic amide forming options are available, isolation costs frequently render these options uneconomic. There is still a pressing need for the development of catalytic, environmentally friendly acylation processes.

Many examples of ester or carbamate formation are covered in the section on protection/deprotection; only those cases where these functional groups survive in the final drug candidate are included here.

Heteroatom alkylations & arylations

This is the largest single class of reactions in this survey, making up 19% of the total. Typically about 90% of drug candidates are N-containing, and an even higher proportion are O-containing.[12] When disconnecting drug candidates, it is thus not surprising that alkylation and arylation at nitrogen and oxygen emerge as major reactions in the construction of the molecules. Table 10 shows the breakdown of *N*-, *O*- and *S*-alkylations.

The synthesis of sibenadet hydrochloride, Scheme 6, shows a molecule constructed by *O*-, *N*- and *S*-alkylations.[14] A phase transfer alkylation of an alcohol is used to construct the O–C bond, a free radical addition to an olefin is used in the construction of the S–C bond and a conjugate addition of an amine to a vinyl sulfone is used in the formation of the N–C bond. It should be noted that the latter two addition reactions are highly atom efficient. The free radical thiol addition could be carried out neat on the laboratory scale, but had to be diluted with toluene for scale up, as it was a very exothermic reaction.



Scheme 6 *Reagents and Conditions*: i) allyl bromide, NaOH, Bu_4NHSO_4, Tol, H_2O, 97%. ii) $HSCH_2CH_2OH$, AIBN, Tol, 90%. iii) NEt_3, MeOH then HCl 75%.

N-Substitution is typically achieved by one of 3 strategies: (i) direct reaction with alkyl-X or aryl-X, (ii) reductive alkylation using an aldehyde or ketone, or (iii) acylation plus reduction of the carbonyl. The acylation–reduction strategy avoids the need to handle alkylating agents and would be more widely used if safer (preferably catalytic) bulk amide reduction methods were developed. This section deals only with the first 2 strategies, as acylation and reduction are covered separately. Table 11 summarises the breakdown of *N*-substitution reactions in this survey.

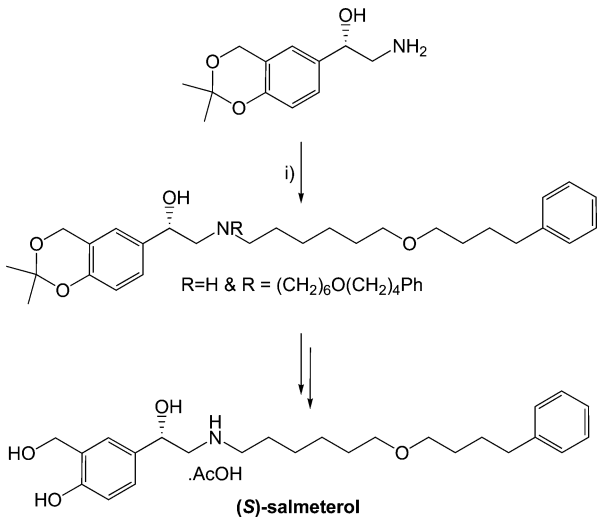
S_N2 reactions are still widely used, though the stringent requirements to ensure that only vanishingly small levels of residual alkylating agents are present in drug candidates, discourages such substitutions late in a synthesis. The synthesis of (***S***)-**salmeterol**, Scheme 7, highlights the difficulties in controlling S_N2 style *N*-alkylation where overalkylation can be a problem.[15] The reductive *N*-alkylations (reductive aminations) in Table 11 are 1-pot conversions; additional cases where the imine has been pre-formed are to be found in the reduction section. Nucleophilic aromatic substitution has long been used in pharmaceutical syntheses for electron deficient aromatics. However, the development of Pd-mediated aryl C–N bond forming methodology, led by Buchwald and Hartwig,[16] is an example of where a technology advance has

Table 10 Categories of alkylation/arylation

Heteroatom	Number	Frequency
N-Substitution	112	57%
O-Substitution	54	28%
S-Substitution	16	8%
Other	14	7%
Total	196	

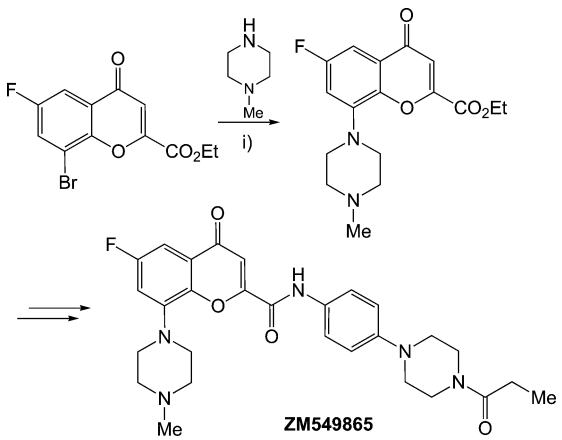
Table 11 *N*-Substitution reactions

	Number	Frequency
N-Alkylation with Alk-X	40	36%
Reductive *N*-alkylation	22	20%
N-Arylation with Ar–X	19	17%
Amide *N*-alkylation	11	10%
Aniline *N*-alkylation	11	10%
Heteroaryl *N*-alkylation	9	8%
Total	125	



Scheme 7 *Reagents and conditions*: i) $Br(CH_2)_6O(CH_2)_4Ph$, DMF, 70%.

led to an increase in the use of a transformation in bulk syntheses, see Scheme 8 for the preparation of **ZM549865**.[17]



Scheme 8 *Reagents and conditions*: i) $Pd_2(dba)_3$, *rac*-BINAP, Cs_2CO_3, anisole, 73%.

Alkylations of phenols (30 instances) and alcohols (24 instances) to give ethers were encountered with moderate frequency, while alkylations at sulfur (16 in total) were less common. Two examples of the Mitsunobu reaction for heteroatom substitution occurred in the database, but this reaction is not often used on scale because of the thermal hazards associated with azodicarboxylates.

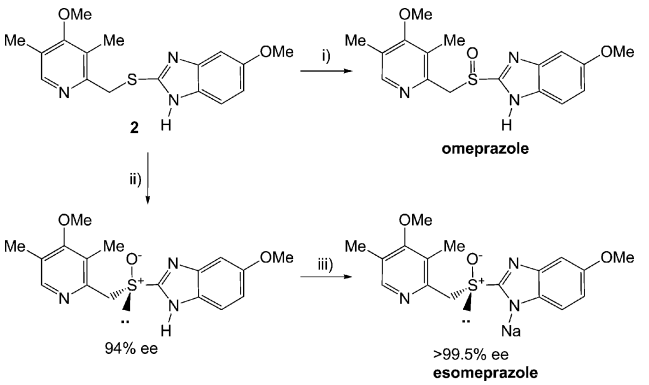
Oxidation reactions

The use of oxidation reactions in the preparation of candidate drug molecules is low, only 3.9% of all reactions covered in this survey. A breakdown of the oxidation reactions used is given in Table 12.

The most common oxidation reaction is oxidation of sulfide to sulfoxide or sulfone. Scheme 9 shows both the racemic stoichiometric oxidation and the more recent enantio-controlled catalytic oxidation of a common intermediate **2** for **omeprazole** and **esomeprazole**.[18]

Table 12 Oxidations

	Number	Frequency
Oxidation at sulfur	10	25%
Alcohol oxidation	7	17%
Oxidation at nitrogen	6	15%
Alkene oxidative cleavage	6	15%
Benzylic/allylic oxidation	6	15%
Alkene oxidation	2	5%
Other	3	8%
Total	40	



Scheme 9 *Reagents and conditions*: i) mCPBA, CH_2Cl_2, 92%. ii) $Ti(O\text{-}iPr)_4$, (*S*,*S*)-diethyl tartrate, cumene hydroperoxide, *N*,*N*-diisopropylethylamine, Tol, H_2O. iii) NaOH, acetonitrile, H_2O, 62%.

Adjustment of the oxidation state from alcohol to carbonyl is rarely performed even though it is a common feature of many published target-orientated syntheses. The problems associated with many oxidation methods probably mean that syntheses requiring them are avoided if possible. The heavy metals used in most oxidation reactions can cause problems in removal and must only be present in trace amounts in the final product (<10 ppm). Many oxidising agents are high energy species, giving rise to thermal hazards and a lack of chemo-selectivity.

Reduction reactions

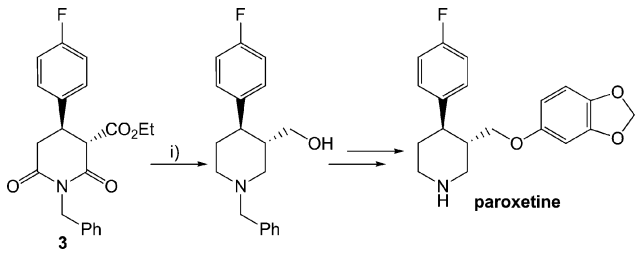
Reduction reactions are used much more frequently than oxidative transformations totalling 9% of reactions (14% if reductive aminations and removal of benzyl-type protecting groups are included). The main reductive transformations are summarised in Table 13.

Catalytic hydrogenation over precious metal catalysts is the most frequently used technique (47%), followed by hydride (32%) and borane reduction (10%). Reductive amination is a frequently used technique to form carbon–nitrogen bonds and is usually carried out with sodium cyanoborohydride or sodium triacetoxyborohydride initially, but these reagents are often superseded by catalytic hydrogenation over an appropriate catalyst on the industrial scale.

It is striking that none of the reductions of carboxylic acid derivatives employ catalytic hydrogenations. The global reduction of the **paroxetine** intermediate **3**, Scheme 10, using excess $LiAlH_4$ exemplifies this type of approach.[19] Hydride and borane reagents are hazardous[20] and lead to both complex work-up procedures and high levels of waste.

Table 13 Reduction reactions

	Heterogeneous Pd/Pt/Ni, H_2	Homogeneous, H_2	Hydride; $LiAlH_4$, $NaBH_4$, DIBALH	Borane	Other	Total
Imine/nitrile to amine	7	—	5	1	1	14
NO_x to NH_2	16	—	—	—	5	21
Alkene to alkane	6	4	2	—	—	12
Alkyne to alkane	1	—	—	—	—	1
Amide to amine	—	—	3	3	—	6
Amide to aldehyde	—	—	2	—	—	2
Ester to alcohol	—	—	8	2	—	10
Ketone to alcohol	—	—	4	3	—	7
Aryl/heteroaryl to fully saturated	6	—	—	—	—	6
Other	5	—	5	1	4	15
Total	41	4	29	10	10	94



Scheme 10 *Reagents and conditions*: i) $LiAlH_4$, THF, 90%.

Catalytic hydrogenation using H_2 gas is the most atom efficient process and developing such methods for all classes of reductions would be of benefit.

C–C bond forming reactions

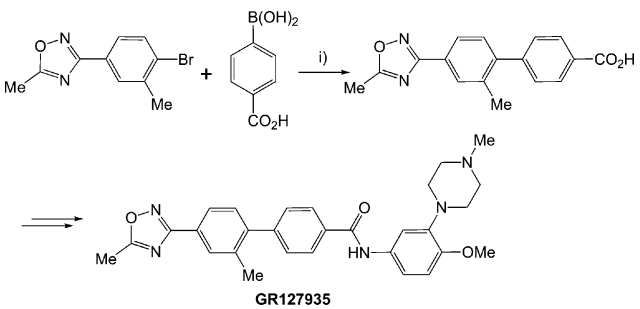
The main C–C bond forming reactions used are shown in Table 14.

The transformations listed in the table are those that make up >10% of the C–C bond forming reactions. This makes up *ca.* 50% of the total, which shows that much of this category does not fit into any significant trend.

The importance of palladium chemistry is clearly evident from the table, although its use is probably more prevalent in early rather than late development. The largest reaction type within the Pd reactions is the Suzuki reaction. Its popularity is derived from the easy accessibility of the two components, convenient reaction conditions, broad functional group tolerance and easy removal of the inorganic by-product, although Pd removal can be problematic.[21] The boronate used in the Suzuki coupling is often formed using an organometallic derived from a halide, linking the increased use of the Suzuki reaction to the improved understanding and operation of organometallic chemistry. Organometallics used directly in C–C bond forming make up an important group of reactions themselves. The advances in the understanding and operation of these reactions have made scale-up almost routine and can be carried out with confidence. The synthesis of **GR127935**, Scheme 11, shows the use of a palladium catalysed Suzuki reaction. The development route moved its position from the final to the penultimate step in the synthesis to control the levels of residual palladium in the final product.[1]

Table 14 C–C bond forming reactions

	Number	Frequency
Pd Catalysis	26	22%
Suzuki	*13*	
Heck	*7*	
Ester condensation	16	14%
Organometallic	14	12%
Aryl–Met	*9*	
Directed lithiation	*3*	
Grignard	*2*	
Friedel–Crafts	12	10%
Other	48	41%
Total	116	



Scheme 11 *Reagents and conditions*: i) Pd–C, Na_2CO_3, *aq* DME, 70%.

It is probably not surprising that ester condensation and Friedel–Crafts reactions still make up a large proportion of the C–C bond forming reactions used. They are well-established reactions where the difficulties are well understood and the possible side-reactions identified.

It is interesting to note that, despite the huge academic effort invested in asymmetric C–C bond forming reactions, they hardly appear in the list of reactions used in process development.

Functional group interconversions (FGIs)

The main FGIs (top five) are shown in Table 15.

The three largest categories in this class of reactions (acid to acid chloride, hydroxyl to halide and amide to imidoyl chloride) are directly related to main reactions used for molecular construction, acylation and heteroatom alkylation/arylation. The acid chloride figure is understated because where the acid chloride is used *in situ* it has been classified with the acylations. These kinds of activations for further reaction would be recognisable to a chemist from a 100 years ago. Many FGIs are really functional group activations,

Table 15 Functional group interconversions

FGI	Number	Frequency
Alcohol to halide[a]	23	22%
Amide to imidoyl chloride	15	14%
Acid to acid chloride	14	13%
Dehydrations	10	10%
Carbonyl to C=NX	8	8%
Other	34	33%
Total	104	

[a] Includes sulfonate esters.

i.e. are performed to enable construction. These transformations would become redundant if methods for the direct conversion of acids to amides, alcohols to amines *etc.* were devised.

Functional group additions (FGAs)

The main functional group additions are shown in the following Table 16 and form only 3.2% of the total. There is a general preference for purchasing starting materials that contain some functionality rather than introducing it. This is due to the problems of selectivity when introducing functional groups and it is preferable to control the quality of the starting materials. Impurities from overhalogenation are usually easier to remove in small molecules by distillation rather than in larger molecules by crystallisation.

Although some of the halogenation will be retained in the final compound, the nitro group is rarely retained in the drug candidates and sulfonations are usually followed by conversion to sulfonamides.

Fluorine is common in many drug substances but there are no examples of fluorination within this data set. The difficulty of introducing it to any complex substrate means the process development is heavily dependent on specialist fine chemical manufacturers for the supply of starting materials, see Fig. 4.

Conclusions

Construction

It is clear from the data that modifying reactions still make up a large part of the chemical transformations. To increase chemical efficiency, improvements in chemoselectivity are still required.

In the design of syntheses there is still a heavy reliance on the use of alkylating agents, especially for C–N formation. Stringent quality requirements to reduce genotoxic impurities in the final product mean that the use of alkylating agents is becoming increasingly problematic. A method for alkylating an amine with, for example, an alcohol would be of great potential benefit.

Table 16 Functional group additions

FGA	Number	Frequency
Halogenation	17	52%
Nitration	12	36%
Sulfonation	2	6%
Other	2	6%
Total	33	

Palladium catalysed reactions have become prevalent in recent years for the formation of C–C and C–N bonds. Their use in scale-up to manufacture still requires a large experimental effort to identify the appropriate catalyst, ligand, solvent and reaction conditions. Having done this, reproducibility and robustness (*e.g.* with regard to substrate quality, small variations in variables) can remain a problem.

A common synthetic sequence is the conversion of an ester to an acid, activation of the acid and conversion to the amide. These are three synthetic steps to accomplish one transformation. There are a large number of methods for the conversion of the acid to the amide, but they require multiple stoichiometric components to effect a simple dehydration.

Chirality

Resolutions remain an important and cost-effective approach to chiral molecules, especially if carried out early in the synthesis. Efficient construction followed by resolution is usually more effective than designing a synthesis around asymmetric technology. Asymmetric transformations are more often carried out on small molecules by the fine chemicals industry than on drug-like molecules late in the synthesis.

Redox

Reductions are far more common than oxidations. Catalytic hydrogenation is highly atom efficient and is well developed for several transformations. However, it is strikingly absent as a technique for reductions of amides and esters, where stoichiometric hydride reagents are the norm despite the difficulties of using these reagents. The development of a mild catalytic hydrogenation to effect these transformations would find widespread industrial application.

An important and interesting addition to reduction technology would be the transformation of an ester to an ether.

In contrast to reductions, there are relatively few atom efficient, chemoselective and environmentally acceptable oxidation methods. As a consequence, oxidations are often designed out of syntheses. The discovery of new chemoselective oxidations, particularly if catalytic, would greatly increase flexibility in synthetic design.

FGI and FGA

The importance of FGA is understated by the survey because the purchased starting materials tend to incorporate the required functionality. A survey of the FGAs used in the fine chemicals industry would probably be more instructive. For example, fluoro substituted aromatics or trifluoromethyl groups are widespread but are nearly always present in the starting materials.

Protections and deprotections

Protections occur infrequently in the data set because they are usually incorporated in the starting materials. Deprotection usually occurs at a late stage, where minimising impurity formation is of paramount importance and by-products from the deprotection must be benign.

Environmental impact

The use of solvents has not been discussed in this review, as this a survey on transformations, but it should be remembered that solvent use is generally the biggest contributor to waste in pharmaceutical production.[22] Recent alternative media such as ionic liquids and supercritical fluids have had little impact to date[23] compared to recycling, using single solvent systems and avoidance of chlorinated solvents.[24]

Acylations are one of the most common transformations but are generally atom inefficient processes. Development of catalytic, low waste acylation methods would significantly improve the environmental performance of many syntheses.

Acknowledgements

The authors would like to thank: Norman Lewis, Gary Kelly (GlaxoSmithKline), David B. Ripin and Nick Thomson (Pfizer) for their assistance with the preliminary research and in collecting the data.

References

1 M. Butters, D. Catterick, A. Craig, A. Curzons, D. Dale, A. Gilmore, S. P. Green, I. Marziano, J.-P. Sherlock and W. White, *Chem. Rev.*, 2006, DOI: 10.1021/cr050982w.
2 R. W. Dugger, J. A. Ragan and D. H. Brown Ripin, *Org. Process Res. Dev.*, 2005, **9**, 253–258.
3 W. H. Miller, T. W. Ku, F. E. Ali, W. E. Bondinell, R. R. Calvo, L. D. Davis, K. F. Erhard, L. B. Hall, W. F. Huffman, R. M. Keenan, C. Kwon, K. A. Newlander, S. T. Ross, J. M. Samanen, D. T. Takata and C.-K. Yuan, *Tetrahedron Lett.*, 1995, **36**, 9433–9436; I. P. Andrews, R. J. Atkins, N. F. Badham, R. K. Bellingham, G. F. Breen, J. S. Carey, S. K. Etridge, J. F. Hayes, N. Hussain, D. O. Morgan, A. C. Share, S. A. C. Smith, T. C. Walsgrove and A. S. Wells, *Tetrahedron Lett.*, 2001, **42**, 4915–4917; I. P. Andrews, R. J. Atkins, G. F. Breen, J. S. Carey, M. A. Forth, D. O. Morgan, A. Shamji, A. C. Share, S. A. C. Smith, T. C. Walsgrove and A. S. Wells, *Org. Process Res. Dev.*, 2003, **7**, 655–662; R. J. Atkins, A. Banks, R. K. Bellingham, G. F. Breen, J. S. Carey, S. K. Etridge, J. F. Hayes, N. Hussain, D. O. Morgan, P. Oxley, S. C. Passey, T. C. Walsgrove and A. S. Wells, *Org. Process Res. Dev.*, 2003, **7**, 663–675.
4 C. M. N. Allerton, J. Blagg, M. E. Bunnage and J. Steele, W. O. Pat. Appl., 200214285, 2002; I. Appleby, L. T. Boulton, C. J. Cobley, C. Hill, M. L. Hughes, P. D. de Koning, I. C. Lennon, C. Praquin, J. A. Ramsden, H. J. Samuel and N. Willis, *Org. Lett.*, 2005, **7**, 1931–1934.
5 J. D. Armstrong, III, M. J. Martinelli and C. H. Senanayake, *Tetrahedron: Asymmetry*, 2003, **14**, 3425–3426.
6 D. J. Ager and I. G. Fotheringham, *Curr. Opin. Drug Discovery Dev.*, 2001, **4**, 800–807; I. C. Lennon and P. H. Moran, *Curr. Opin. Drug Discovery Dev.*, 2003, **6**, 855–875; M. Breuer, K. Ditrich, T. Habicher, B. Hauer, M. Kesseler, R. Sturmer and T. Zelinski, *Angew. Chem., Int. Ed.*, 2004, **43**, 788–824.
7 U. C. Dyer, D. A. Henderson and M. B. Mitchell, *Org. Process Res. Dev.*, 1999, **3**, 161–165; J. W. Nieuwenhuijzen, R. F. P. Grimbergen, C. Koopman, R. M. Kellogg, T. R. Vries, K. Pouwer, E. van Echten, B. Kaptein, L. A. Hulshof and Q. B. Broxterman, *Angew. Chem., Int. Ed.*, 2002, **41**, 4281–4286; A. Borghese, V. Libert, T. Zhang and C. A. Alt, *Org. Process Res. Dev.*, 2004, **8**, 532–534.
8 M. Butters, J. Ebbs, S. P. Green, J. MacRae, M. C. Morland, C. W. Murtiashaw and A. J. Pettman, *Org. Process Res. Dev.*, 2001, **5**, 28–36.
9 A. E. Brandstrom and B. R. Lamm, Eur. Pat. Appl., 103553, 1984; S.-Y. Chou and S.-F. Chen, *Heterocycles*, 1997, **45**, 77–85.
10 A. Matsushita, M. Oda, Y. Kawachi and J. Chika W. O. Pat. Appl., 2003006439; K. Yamamoto, Y. G. Chen and F. G. Buono, *Org. Lett.*, 2005, **7**, 4673–4676.
11 T. W. Greene and P. G. M. Wuts, *Protecting Groups in Organic Synthesis*, Wiley, New York, 3rd edn, 1999.
12 *Med Ad News*, Top 200 World's Best Selling Medicines, **23**, ((5)), 2004, pp. 60–64.
13 P. J. Dunn, W. Hoffmann, Y. Kang, J. C. Mitchell and M. J. Snowden, *Org. Process Res. Dev.*, 2005, **9**, 956–961.
14 M. E. Giles, C. Thomson, S. C. Eyley, A. J. Cole, C. J. Goodwin, P. A. Hurved, A. J. G. Morlin, J. Tornos, S. Atkinson, C. Just, J. C. Dean, J. T. Singleton, A. J. Longton, I. Woodland, A. Teasdale, B. Gregertsen, H. Else, M. S. Athwal, S. Tatterton, J. M. Knott, N. Thompson and S. J. Smith, *Org. Process Res. Dev.*, 2004, **8**, 628–642.
15 P. A. Procopiou, G. E. Morton, M. Todd and G. Webb, *Tetrahedron: Asymmetry*, 2001, **12**, 2005–2008.
16 J. F. Hartwig, in *Handbook of Organopalladium Chemistry for Organic Synthesis*, ed. E. I. Negishi, Wiley-Interscince, New York, 2002, vol. 1, p. 1051; A. R. Muci and S. L. Buchwald, *Top. Curr. Chem.*, 2002, **219**, 131–209; S. L. Buchwald, C. Mauger, G. Migani and U. Scholz, *Adv. Synth. Catal.*, 2006, **348**, 23–39.
17 G. E. Robinson, O. R. Cunningham, M. Dekhane, J. C. McManus, A. O'Kearney-McMullan, A. M. Mirajkar, V. Mishra, A. K. Norton, B. Venugopalan and E. G. Williams, *Org. Process Res. Dev.*, 2004, **8**, 925–930.
18 H. Cotton, T. Elebring, M. Larsson, L. Li, H. Sorensen and S. von Unge, *Tetrahedron: Asymmetry*, 2000, **11**, 3819–3825.
19 D. A. Greenhalgh and N. S. Simpkins, *Synlett*, 2002, **12**, 2074–2076; C. D. Gill, D. A. Greenhalgh and N. S. Simpkins, *Tetrahedron*, 2003, **59**, 9213–9230.
20 *Bretherick's Handbook of Reactive Chemical Hazards*, ed. P. Urben, 6th edn, 1999, vol 2, p. 58.
21 C. E. Garrett and K. Prasad, *Adv. Synth. Catal.*, 2004, **346**, 889–900; J. T. Bien, G. C. Lane and M. R. Oberholzer, *Top. Organomet. Chem.*, 2004, **6**, 263–283.
22 D. J. C. Constable, A. D. Curzons and V. L. Cunningham, *Green Chem.*, 2002, **4**, 521–527.
23 R. A. Sheldon, *Green Chem.*, 2005, **7**, 267–278.
24 D. H. Brown Ripin and M. Vetelino, *Synlett*, 2003, **15**, 2353.