Key Organics
Chemistry | Innovation | Quality

Technical Note

Compound Quality in Medicinal Chemistry

Quality – the impact of molecular properties

The concept of improving the quality of the compounds we screen in drug discovery projects is one which is likely to be familiar to most, if not all, medicinal chemists. A key unstated assumption is that the term 'quality' refers to the **molecular properties** of the compound, and how they relate to enhancing the likelihood of the compound becoming a drug. Often this important concept is synonymous with the application of Lipinski's rules or related concepts depending on the target indication, discovery modality employed, or mode of delivery (Table 1).

Lipinski Rules ("Rule of 5")	Fragment Rules ("Rule of 3")	CNS Penetrant API
MW ≤ 500	MW ≤ 300	MW ≤ 400
cLogP ≤ 5	cLogP ≤ 3	cLogP ≤ 5
HBA ≤ 10	HBD ≤ 3	HBD ≤ 3
HBD ≤ 5	HBA ≤ 3	HBA ≤ 7
# RB < 10		# RB < 8

Table 1: *Example compound 'rule sets' (MW = molecular weight / Da, HBA = hydrogen bond acceptors, HBD = hydrogen bond donors, # RB = number of rotatable bonds)*

In recent years the molecular real-estate occupied by new and emerging drugs has been changing, and given medicinal chemists reason to pause and think carefully about the design parameters they employ. The emergence of drugs designed to disrupt protein-protein interactions (e.g. venetoclax, approved 2016, Figure 1), or to leverage targeted protein degradation (e.g. ARV-110) has necessitated the design of compounds which sit well outside conventional 'drug space', as defined by existing rule sets.

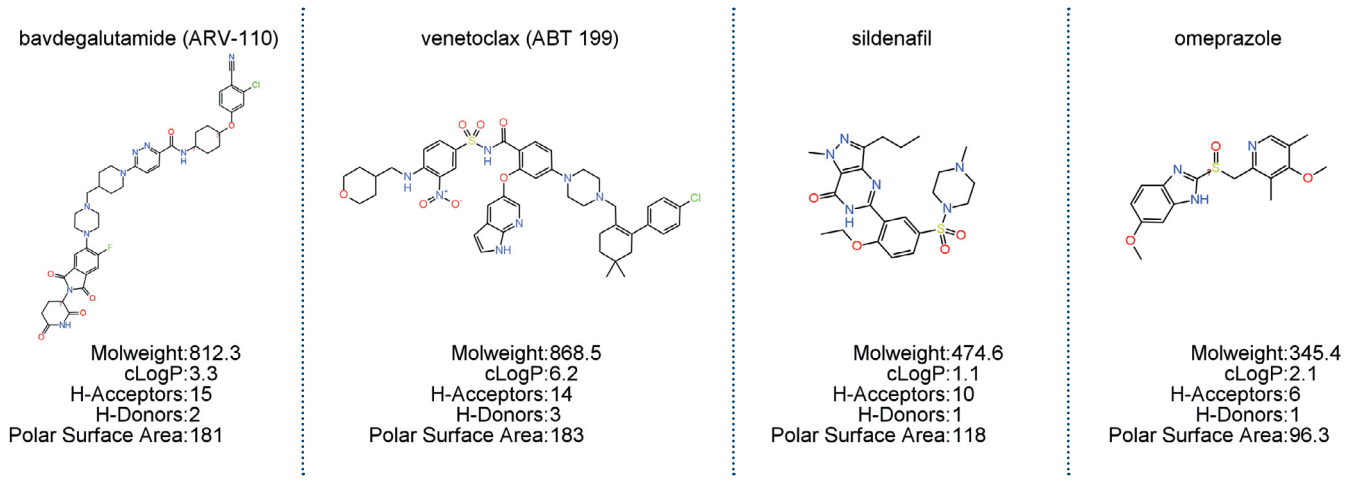

Figure 1: *The evolution of small molecule property space*

With an increasing number of compounds and programs focused on these newer discovery strategies, alongside emerging new classes of therapeutic, for example, drugs targeting non-coding RNA, this trend of shifting property space is likely to continue. It is even possible that the term 'small molecule' may require rethinking or redefining. Perhaps it may even fall out of use altogether as the space between conventional small molecules and 'large(r)' molecules (protein, peptides, nucleic acids etc.) becomes ever more blurred (Figure 2).

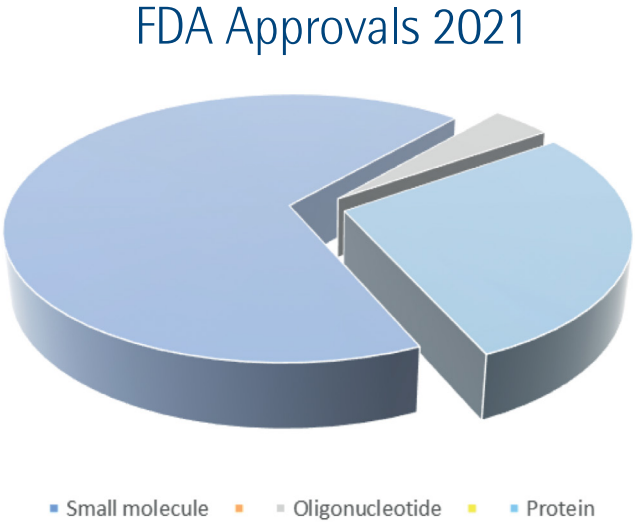

Figure 2:
FDA approvals 2021, by drug modality

Quality – looking beyond molecular properties to compound provenance

At the heart of all of this lies a subject which is rarely covered in any detail, or even documented in the medicinal chemistry literature. That is the concept of 'quality' as it relates to the **provenance of the compound being tested**, be that a fragment, a conventional Lipinski-compliant lead-like compound, or a newer PROTAC-type degrader. All compounds tested in biological assays are the products of one or more sequential chemical reactions, and therefore originate from a synthetic sequence of some sort. By definition, a chemical reaction has inputs (e.g. reagents, starting materials) and outputs (e.g. desired products, side products, reagent by-products) along with solvents, reaction vessels and the like which facilitate the desired transformation, but are intended (or expected) to remain unchanged throughout. Scale-up / process and manufacturing chemists are well accustomed to deeper consideration of these aspects, but some factors may be thought about less frequently by chemists working on earlier stage compounds such as those evaluated during hit identification and lead optimisation.

Chemistry teams nowadays are fortunate to have access to a wide range of tools to facilitate the design and synthesis of compounds, be they novel or previously described. This enables individuals and teams to select (for example) the best starting materials (most cost / time effective) along with the reaction conditions which are most likely to deliver a good yield of desired product. Due to the sheer volume of compounds which have been described from both peer reviewed and non-peer reviewed sources, there is also a vast quantity of analytical and characterisation data available to refer to or provide comparisons for identification purposes. For example, the Chemical Abstracts Service (CAS) references over 190 million organic and inorganic substances in total, with thousands of new ones being added daily. In terms of purchase, free-to-access databases such as ZINC15 contain over 230 million compounds, and e-Molecules comprises 24.6 million screening compounds and 9 million building blocks.

However, the magnitude of this data set is also where a potential problem arises for chemists and is perhaps in part attributable to 'information overload'. The assumption that may often be made when using this plethora of chemical reaction and compound-based data in decision making for to give a particular compound output (O) *i.e.* selecting best compound inputs (I) and suitable chemical reactions (R) is that all information, individual data points and parameters used are fit-for-purpose for the end user (see Figure 3 for a simple process).

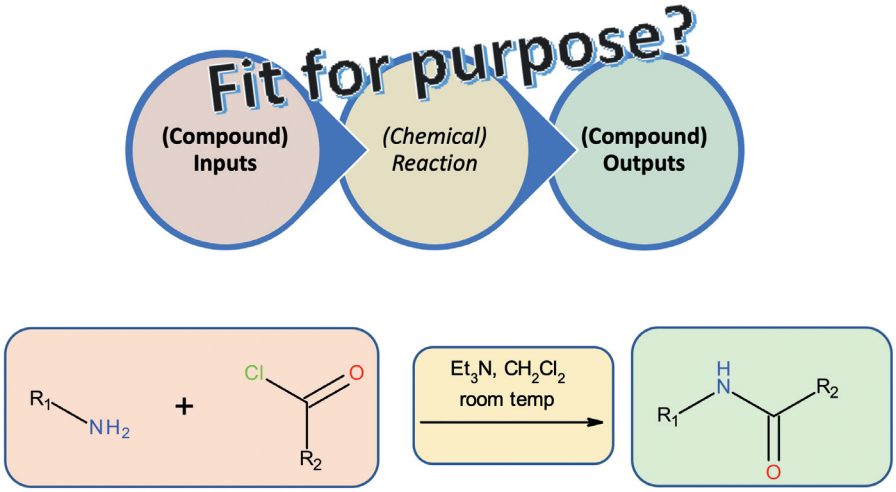

Figure 3: *A simple chemical reaction classified into basic **I**nputs, **R**eaction, **O**utputs*

Whilst the majority of the time these assumptions may be correct (and ignoring them therefore has no significant consequence), from time-to-time one or more key assumptions may be erroneous. For example, the wrong building block (input) may be selected, or the chemical reaction used may be sub-optimal, for reasons which are not immediately obvious. Chemists usually anticipate this type of random error occurring from time-to-time; it's put down to experiences, and and they use one of several possible back-up options to ensure success in obtaining the desired compound output. However, the situation is potentially more complex than this, with not all assumptions always being factored into the decision-making process. The end result then can be negative consequences including wasted time, financial loss and even possible pursing of the wrong direction for a project. In some of the worst cases these could lead to problems with marketed drugs, those in trials, issues with reproducibility or retractions of papers. With this in mind it could be viewed as preferable to generate data to confirm an assumption (i.e. use 'data not dogma'), even if the answer may seem obvious, should the potential impact of being wrong having detrimental consequences for a project.

Looking at this more specifically, a (non-exhaustive) selection of additional variables, and common, assumptions made about them, potential consequence(s) and possible solution(s) are shown below (Table 1, categorised into Input / Reaction / Output). Recently, several papers have been published (hyperlinked) which have questioned one or more of the various factors that take place at each point in the process of preparing a compound.

	Variable	Potential consequence(s)	Possible solution(s)
Input	Is the reagent supplied as stated on the bottle label? Could it have been labelled wrongly, or become contaminated by a previous user?	• Failed reaction • Wrong product prepared / isolated • Safety issue	• Check data sheet • Independent analysis at point of receipt / use
	Isomers (region, geometric etc) of the reagent are as stated on the bottle / vial	• Failed reaction • Wrong product prepared / isolated	• Independent analysis at point of receipt / use
	How long has the bottle of reagent sat on the shelf in the storeroom, and under what conditions?	• Failed reaction • Unexpected low yield	• Stock room logs • Independent analysis pre-use
Reactions	Are the reaction conditions described in a paper as producing 85% yield of product correct, or was it a 'one-off' best case?	• Lower yield than stated • Loss of starting material • Failed reaction	• Adopt same approach as biology with 'n=3' technical replicates
	Are all components stable under stated conditions?	• Failed reaction • Failure to isolate desired product	• Careful monitoring
Output	Has structure been assigned accurately by paper / patent authors and/or has all necessary analytical data been used, and interpreted correctly?	• Wrong compound tested • Inaccurate SAR • Project direction changed	• Independent confirmation of structure / regiochemistry using in-house data with no inherent assumptions
	Is the isolated product a free base, salt, hydrate or combination thereof (despite what may be stated in a procedure)	• Unexpected physicochemical properties • Incorrect stoichiometry calculations	• Enhanced analysis, (e.g.) elemental / combustion
	How sure can one be that there are no undesired, low level, and potentially problem-causing impurities in any 'purified' and isolated sample(s)?	• Failed reaction • 'Surprise' reaction(s) • Unexpected or erroneous screening results	• Use enhanced purification and more stringent / targeted analysis

The list above is far from comprehensive but is intended to provide some food for thought and illustrate common assumptions made during the synthetic process about the wide range of variables, the possible pitfalls resulting from them, and the generally simple solutions to mitigate them.

The main 'fit for purpose' questions which medicinal chemists should ask themselves before starting, during and after any synthetic process (IRO) are:

1. INPUT: *verify reagent identity and quality*
2. REACTION: *verify procedure suitability and analytical data quality*
3. OUTPUT: *verify product identity and quality*