Advanced Drug Delivery Reviews 101 (2016) 34–41

Contents lists available at ScienceDirect

Advanced Drug Delivery Reviews

journal homepage: www.elsevier.com/locate/addr

Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions☆

Christopher A. Lipinski *

10 Connshire Drive, Waterford, CT 06385-4122, USA

ARTICLE INFO

Article history:
Received 29 December 2015
Received in revised form 22 April 2016
Accepted 27 April 2016
Available online 3 May 2016

Keywords:
Biologically active chemistry space
Binding pockets in folded proteins
Protein binding pocket evolution
Characteristics of protein ligands
Natural product chemistry synthons
Shape change in natural products
H-bonding in natural products
Navitoclax a rule of 5 outlier

ABSTRACT

The rule of five (Ro5), based on physicochemical profiles of phase II drugs, is consistent with structural limitations in protein targets and the drug target ligands. Three of four parameters in Ro5 are fundamental to the structure of both target and drug binding sites. The chemical structure of the drug ligand depends on the ligand chemistry and design philosophy. Two extremes of chemical structure and design philosophy exist; ligands constructed in the medicinal chemistry synthesis laboratory without input from natural selection and natural product (NP) metabolites biosynthesized based on evolutionary selection. Exceptions to Ro5 are found mostly among NPs. Chemistry chameleon-like behavior of some NPs due to intra-molecular hydrogen bonding as exemplified by cyclosporine A is a strong contributor to NP Ro5 outliers. The fragment derived, drug Navitoclax is an example of the extensive expertise, resources, time and key decisions required for the rare discovery of a non-NP Ro5 outlier.

Contents

☆ This review is part of the *Advanced Drug Delivery Reviews* theme issue on "Understanding the challenges of beyond-rule-of-5 compounds".

* Tel.: +1 860 326 0633.

http://dx.doi.org/10.1016/j.addr.2016.04.029
0169-409X/

1. Introduction

The rule of five (Ro5) was originally published in 1997 and was based on the physicochemical profiles of phase II drugs [1]. In this mini-perspective I reflect on how: The Ro5 is consistent with what we now know about the structural limitations of protein targets and on the drugs that are ligands for the protein targets. Three of the four parameters in Ro5 are fundamental to the structure of both target and drug binding sites. The chemical structure of the drug ligand depends on the chemistry toolkit used in the ligand synthesis and the design philosophy behind the ligand. Two extremes of chemical structure and design philosophy exist; the ligands constructed in the medicinal chemistry synthesis laboratory without input from natural selection and the natural product (NP) metabolites biosynthesized by an organism based on evolutionary selection. It is in this latter, NP class, that one finds most of the exceptions favorable to oral absorption from the Ro5. I hypothesize that the biophysical chemistry chameleon-like behavior of many NPs due to intra-molecular hydrogen bonding as exemplified by cyclosporine A is a strong contributor to Ro5 outliers among NPs. Drug discovery project decisions can occasionally lead to development of a non-NP Ro5 outlier. The orally active anti-cancer drug Navitoclax, now in phase III studies, provides an excellent example of the extensive expertise, resources, time and key decisions required for the rare discovery of a non-NP Ro5 outlier.

1.1. Structural limitations of protein targets and on their ligands the drugs

Based on our current knowledge virtually all of chemical space is devoid of biologically active compounds. “The limits of biologically relevant chemical space are defined by the specific binding interactions between small molecules and the three dimensional molecular recognition patterns on biological molecules, such as proteins, RNA and DNA, which have evolved over billions of years” [2]. Measured in terms of physicochemical properties and topological descriptors of the ligand, therapeutically useful ligands appear to cluster together in galaxies. Clustering of ligand cavities, the binding partner of the ligand, is not well predicted by either sequence or fold space. There are differences in the similarities in sequence-, fold- and cavity space such that cavity space is most similar to ligand binding space [3]. A consequence of this is the cross reactivity observed not only for co-factors (to be expected) but also for ligands of proteins of unrelated sequence or fold (unexpected). In a study of the same ligand binding to unrelated proteins almost half of the instances involved the ligand binding to unrelated residues in the two proteins [4]. The cross reactivity observation is pertinent both to issues of off-target toxicity and to opportunities in drug repurposing. Until the last decade an unanswered question was whether the galaxies of biologically active compounds are evenly and sparsely distributed and therefore hard to find, or whether most of the chemical universe is ‘empty’ (containing no therapeutically interesting compounds), with galaxies of therapeutically interesting compounds scattered far apart [2]. The clear evidence that biologically active compounds exhibit small world clustering behavior [5] indicates that the second alternative is correct, namely that most of chemical space is “empty” and devoid of biologically active compounds and correspondingly that most of protein sequence space is devoid of functional activity [6]. It is important to note some limitations on the observation that biologically active compounds are clustered in protein sequence space. Virtually all the data behind this observation are based on ligands of monomeric proteins. There are too few ligands against protein–protein interactions and too few ligands against nucleotide, carbohydrate or lipid targets to draw any conclusions on the nature of biologically active space against these target classes.

1.2. The immensity of chemical space

The number of possible small molecules is unimaginably large. For example, the chemical space of small molecules with molecular weight of 500 or less and containing the common atoms found in drugs is estimated at 10^{60} [7]. The number of 62 membered peptides containing any of the twenty common amino acids (AA) is approximately10^{80}. The number of atoms in the universe is quoted as 10^{80} [8]. The theoretical chemical space possibilities for any protein likely to be a drug target (i.e. more than 62 AA) is similar or larger than the number of atoms in the universe. How do experimental observation compare to theoretical possibilities?

1.2.1. The number of experimentally observed protein folds

The number of experimentally observed protein folds in the PDB database is just under 1400 and the total of folds is unchanged for almost the last decade [9]. It is highly likely that at about 1400 folds we are at the asymptote of the total number of physically possible protein folds [10]. The shape of the ligand binding pockets in experimental X-ray peptide structures has been computationally estimated and the endogenous ligand binding site is usually the largest, most hydrophobic and geometrically most complex pocket of a protein. Moreover, druggable pockets form a distinct cluster in pocket descriptor space [11].

1.2.2. The total number of pockets

The total number of pockets in a protein is in the range of 1400–2000 depending mostly on whether the pocket is unoccupied by a ligand (which gives the smaller number) or whether the pocket is occupied by a ligand (which gives the larger number). The total also depends somewhat on the definition of shape similarity. The database of pockets occupied by ligands has been termed the “pocketome” [12]. Larger unoccupied pockets can sometimes be occupied by structurally different ligands with movement in the protein to accommodate the ligand with a consequent change in pocket shape. The general pattern is that smaller pockets occupied by smaller ligands tend not to change shape on ligand binding. It is in the larger pockets that pocket shape may change on binding of a larger ligand. For both protein structures (i.e. drug targets) and drug binding sites in the target proteins the small number of shape structural possibilities is an incredibly small fraction of the theoretical possibilities in chemical space.

How does the number of drug target pocket shapes relate to the number of possible shapes for drugs? The number of protein pocket shapes represents a lower limit. The number of drug ligand shapes is certainly much higher than the number of drug target pocket shapes. Many drugs are capable of multiple conformations and the old idea that one should limit the choices by only considering conformations within 3 to 4 kcal of the global minimum is incorrect [13]. There are extreme cases in which the bound conformation of a drug can be as much as 9 kcal above the global minimum energy conformation. In addition, the pocketome compilation is based on protein X-ray structures. Many targets and their ligands have no X-ray data on the target. Finally, a ligand can project out of a protein pocket into solvent space thus increasing the ligand shape possibilities.

1.2.3. The chemical space of proteins and that of drug binding sites

The chemical space of proteins and that of drug binding sites is very limited. Why is this? The answer lies in the biophysics of the close packing of a protein. The crystallization of a protein (i.e. an alpha AA polymer) is determined by an inter play within the protein relating to an achievement of maximum density for the packed protein while still allowing for satisfaction of protein hydrogen bond donors and acceptors. These factors almost always work in opposition to each other and the final state of a protein is the lowest free energy compromise. The same compensatory process works in small molecule crystallization. The driving force leading to maximum density is the result of solvent pressure on the surface of the packed protein. The drive to satisfy all hydrogen bond donors and all the strong and moderate hydrogen bond acceptors is the very large energy penalty resulting from an unsatisfied hydrogen bonding partner. The drive to satisfy hydrogen bonding is the reason for the tightly bound water that is found within a protein structure since there may be insufficient numbers of internal hydrogen bond donors within a protein to satisfy the stronger and moderate hydrogen bond acceptors within the protein. The hydrogen bond donating and accepting ability of tightly bound water ensures that all the important hydrogen bond donors and acceptors in the protein are satisfied. Compared to the well characterized cavity behavior in monomeric proteins, most of the global interaction features in protein–protein interactions has focused on the "hot spots" that might represent a druggable binding site [14]. However as described later in Section 4, the interaction surfaces of protein–protein interactions tend not to have the types of cavities found in monomeric proteins. Nature through evolution has capitalized on the inherent tendency of proteins to interact with one another as a key feature in biological protein–protein interaction signaling networks. The wide range of protein–protein interaction energy possibilities, only some of which are useful in a signaling sense, is a type of emergent phenomenon that is not readily apparent from the relatively limited protein synthesis tool kit of 20 amino acids and post translational modifications.

1.2.4. Compactly packed proteins have inherent biological activity

A compactly packed protein has inherent biological activity that depends on the inevitable cavities and crevices that come from packing imperfections and is independent of any evolutionary consideration. Proteins made in a chemistry synthesis laboratory without any relationship to known biological literature have been shown to have small molecule binding activity [15,16]. Evolution works on what already preexists namely the biological activity inherent in the pockets and crevices of incompletely packed proteins [17]. Current thought is that the shape of the ligand binding site is the fundamental binding site property [18]. Support for this arises from the considerable and perhaps surprising success in the last decade in using shape just by itself as a parameter in medicinal chemistry quantitative structure activity relationships [19]. The tweaking of electrostatic surface properties within a particular cavity shape likely evolved to allow for specificity and greater catalytic activity. There is a limit to the allowable evolutionary change in a cavity. As the cavity evolves toward greater catalytic potency there is a general tendency for the original protein folding to become increasingly unstable. At the stability limit, proteins begin to form structures leading to gradually increasing insoluble aggregates many of which are currently believed to be at the origin of multiple disease pathologies [20]. In addition "the stability of native proteins is primarily determined by hydrophobic interactions between sidechains, while the stability of amyloid fibrils depends more on backbone intermolecular hydrogen bonding interactions" [21]. The protein cavity size distribution modulated by solubility and permeability considerations is largely responsible for the allowable size range of orally acting drugs acting on monomeric target proteins. The cavity size distribution of these targets has been studied and corresponds well with the allowable range of the Ro5 molecular weight parameter [22].

1.2.5. The compactly packed protein model is a bit over simplistic

The compactly packed protein model is a bit over simplistic. In reality, the protein exists as a dynamic equilibrium mixture of protein conformers each of which realizes its minimum Gibbs free energy through different contributions of enthalpy and entropy to the total protein free energy [23]. In X-ray protein structures this dynamic equilibrium manifests as regions of structural disorder, larger atom thermal ellipse boundaries and infrequently as different protein conformations within the same X-ray unit cell. In NMR structures of the smaller proteins suitable for NMR studies the equilibrium is often seen as the direct detection of individual conformers. When X-ray and NMR structures of the same protein are compared the RMSDs of the protein backbones are similar and typically within 1 Å and most of the differences are found in the amino acid side chain positions [24]. This is consistent with the hypothesis that the target cavity shape is determined by the primary sequence and that the electrostatic component of ligand binding is mostly influenced by the amino acid side chain composition.

Several binding modes can accommodate a ligand that is modestly larger than predicted from target site scanning (hot spot scanning) of a monomeric epi-protein. The phenomenon of "induced fit" in which a ligand can induce a larger cavity as in many kinase inhibitors is thought to be due to the capture and stabilization by the ligand of one of the dynamic protein conformers. This phenomenon of activity in a somewhat larger ligand is likely to be more common for those proteins whose evolved function involves some type of significant molecular motion as in a flap moving or a crevice enlarging. A somewhat larger ligand than predicted from epi-protein hot spot scanning can also be accommodated if part of the ligand projects outside of the protein into solvent space. With carefully constructed hydrophilic functionality projecting into solvent space, aqueous solubility of the ligand can be improved by medicinal chemists with little or no loss in overall ligand binding energy.

1.2.6. Hydrogen bonding patterns are critical to the packing of proteins

Hydrogen bonding patterns are critical to the packing of proteins and thus to the formation of the cavities and crevices due to incomplete packing that are the binding sites for drugs. We now know quite a bit about the hydrogen bonding characteristics of drugs. The three dimensional characteristics of hydrogen bonding are well understood from analyses of hydrogen bonds in small molecule complexes in the Cambridge Structural Database and from hydrogen bonds in protein—ligand complexes in the Protein Data Bank. From an analysis of ligands containing multiple hydrogen bond donor groups it is apparent that within a single ligand it is very difficult to accommodate more than about two or three hydrogen bond donor groups with the precise geometry needed for the maximum enthalpic energy benefit arising from the optimal three dimensional positioning of a hydrogen bond [25]. This phenomenon is often discussed in the context of entropy–enthalpy compensation [26]. As a result, as the number of potential hydrogen bond donors in a ligand increases it becomes increasingly likely that they will not contribute in a positive sense to ligand binding and likely will detract from ligand binding. This observation is very consistent with the common medicinal chemistry observation that it is difficult to improve potency by addition of hydrogen bond donor groups. The directionality of hydrogen bonding places restrictions on the number of hydrogen bonds in a ligand quite apart from the effect of hydrogen bond donor and acceptor groups on membrane permeability [27].

1.2.7. The parameters of the Ro5 and their role for protein targets

Three of the four parameters in the Ro5 namely the hydrogen bond donors and acceptors and the molecular weight are fundamental to the structure of both protein target and drug binding sites. It should be noted that the molecular weight is a more easily calculated surrogate for molecular volume or size. Lipophilicity as in the log P parameter in the Ro5 is a composite parameter, namely the ratio of drug solubility in water saturated with n-octanol divided by the drug solubility in n-

octanol saturated with water. The log P parameter is ultimately derived experimentally from an artificial system and thus, unlike the other three Ro5 parameters cannot be directly related to target biophysical properties.

1.3. Natural products (NPs)

The chemical structure of the drug ligand depends on the type of chemistry and the choice of synthetic intermediates used in the ligand synthesis and on the design philosophy behind the ligand. Two extremes of chemical structure and design philosophy can be identified. In the first of these there are the ligands constructed in the medicinal chemistry synthesis laboratory without necessarily any input from the structures of NPs and with input from the range of available chemistry intermediates and the existing repertoire of organic chemistry synthetic transformations. Although it should be noted that, in practice, only a very small subset of known synthetic transformations is typically used to make compounds for biological screening [28]. The second extreme consists of the NPs that are secondary metabolites biosynthesized by an organism based on evolutionary selection. It is in this latter NP class that one finds most of the exceptions to the Ro5 that are favorable to oral absorption.

1.4. Most of the favorable exceptions to Ro5 occur among NPs

Most of the favorable exceptions to Ro5 occur among NPs. As a medicinal chemist discussing NPs, I intend to focus on the chemistry and evolutionary aspects and how these relate to drug discovery in general and the Ro5 in particular. The NP chemistry assembly toolkit is "limited" when compared to the much broader set of synthetic transformations available in current synthetic organic chemistry. Among the NPs there are only about eight building blocks; namely C1, C2, C5, C6C3 (phenylpropyl), C6C2N (phenethylamine), indole, C4N (pyrrolidine), and C5N (piperidine) [29]. I think it is fair to say that no modern medicinal chemist would ever limit themselves to such a limited range of synthons. One can speculate as to the cause of this limited set of basic carbon building blocks. The current set is certainly capable of producing an extraordinarily large number of final products because of the advantages inherent in biosynthetic combinatorial chemistry and is self-evidently adequate in an evolutionary sense. One could conservatively conclude that the patterns of NP assembly are a triumph of millennia of biology evolutionary selection and screening that completely outweighs the theoretical advantage of the larger experimental chemistry synthesis toolkit developed largely over the last 100 years. The current total of small molecule organic single molecule synthetic compounds ever made totals about 70 million in the Chemical Abstracts Service database and about the same number in the public domain UniChem database [30]. Only a fraction of these have ever been tested for biological activity with the majority of biological testing occurring since the advent of high throughput screening in the early 1990s.

1.5. NP synthons are related to secondary metabolic target

A more speculative view regarding the patterns of NP assembly relates to possible inherent advantages of NP synthons as they relate to secondary metabolite target tissue penetration. It seems to me that a secondary metabolite is of limited or no evolutionary advantage if it cannot attain its intended target. The ability of some secondary metabolites to self-modulate biophysical properties related to membrane and target penetration could have been an important feedback in the evolution of a combinatorial biosynthetic pathway and to the assembly and chemistry fine tuning of NP synthons. I speculate that the chameleon like property of NP's like cyclosporine-A (vide supra) might mostly be available in the type of large macrocycle and substituent chemistry found in a few NPs and not in the overwhelming majority of the near 70 million organic synthetics made to date. To be fair, there are some purely synthetic approaches to small molecule chemistry that in a template sense mimic the structure of regular secondary structure elements, such as alpha helices, beta strands and reverse turns found in many enzyme-substrate/inhibitor and receptor agonist/antagonist complexes [31].

The synthons for NPs derive from synthons in primary metabolism and in the current way a medicinal chemist thinks are not terribly efficient. For example, to biosynthesize shikimic acid, the precursor to an aromatic ring from a carbohydrate primary metabolism synthon takes seven steps with additional steps toward the human essential aromatic amino acids phenylalanine, tyrosine and tryptophan.

1.6. The repertoire of chemistry synthetic transformations is huge

The repertoire of chemistry synthetic transformations is huge, standing in contrast to the limited range of NP synthons. For example the 6th edition of March's Advanced Organic Chemistry consists of 2356 pages packed with synthetic chemical transformations, mechanisms for these and discussions of scope and limitations. Wikipedia as of March 21, 2016 lists 723 organic reaction types. In considering NP secondary metabolites the analogy to combinatorial chemistry is perhaps apt since recent evidence suggests that only a fraction of the NP secondary metabolites capable of biosynthesis have actually been tested by and found active through evolution [32].

1.7. The chemistry tool kit for NP biosynthesis might be special

The chemistry tool kit for NP biosynthesis might be special apart from the millennia long NP evolutionary selection process. This is a subject of some debate best illustrated by the discussion of the flavonoids. Flavonoids are famous, perhaps notorious, for their ability to bind to multiple protein targets. The argument is made that since flavonoids are biosynthesized by a protein they are better suited for affinity to protein targets [33,34].

1.8. Cheminformatics to relate chemical structure to biological activity

A cheminformatics classification of the chemical structure based on the underlying chemical scaffold and the use of chemical similarity calculations among purely synthetic drugs allows successful analysis of the relationship of chemical structure to biological activity. There are literally dozens of publications on this general approach. The cheminformatics approach works among laboratory made synthetic drugs because, in general, the underlying scaffold is fairly rigid and so there is a one to one equivalence between scaffold structure and chemical shape/polarity signature.

The cheminformatics classification of the chemical structure based on the underlying chemical scaffold has been successfully applied to the most abundant (and often conformationally rigid) natural products like steroids or flavones [35] as well as across databases of natural products [36] but there is little information on the numerically much smaller sets of macrocyclic and theoretically conformationally flexible natural products (e.g. ketolides). I note that the importance of macrolide conformational flexibility to the efficacy (as opposed to permeability) of macrolides as beyond Ro5 ligands is controversial with a recent review [37] supporting the position that it is the edge on and face on binding modes of macrocycles rather than conformational flexibility per se that account for the prevalence of macrocycles in beyond Ro5 compounds.

1.9. NPs can be conformationally flexible

NPs can exist as potentially conformationally flexible ring structures or as potentially conformationally flexible acyclic structures and both of these classes can form one (or more) intramolecular hydrogen bond (HB) between an HB acceptor like a carbonyl group and hydrogen

from an HB donor like an amide N–H or an O–H. The energetics of an intra-molecular HB in aqueous environment is currently somewhat difficult to calculate and the calculation errors for a compound capable of multiple hydrogen bonding possibilities are large enough so as to make somewhat problematical any firm conclusions about the structure of the most stable intramolecular HB conformer. In the literature one can find schemes for classifying NPs as to likely biosynthetic origin and also by scaffold [36]. However, one cannot find an atlas or compendium of the actual shapes and polarities in aqueous medium of conformationally flexible intramolecularly hydrogen bonded natural products. This scientific deficiency in understanding that I have just described does not mean that conformationally flexible intramolecularly hydrogen bonded NPs cannot be used to attain a biological goal. This is where evolution comes in. A billion or more years (or less) of empirical screening for a desired biological phenotypic effect from a natural product secondary metabolite can be very successful indeed without any understanding whatsoever of what is going on in a biophysical or mechanistic sense. It should also be noted that the phenotypic evolutionary selection of NPs does not mean that the NP is selective for a particular mechanism. In fact there is considerable speculation that evolutionary selection may often drive for polypharmacology and that a desired phenotypic biological effect may often be due to a mixture of often structurally related NPs [38].

1.10. NP's interact with target nodes of higher biological network connectivity

NPs interact with target nodes of higher biological network connectivity than are typically the targets of most synthetic drugs [39]. This is called the "central hit strategy" in that the drug selectively targets central node/edges of the flexible networks of infectious agents or cancer cells to kill them [40]. The "network influence strategy" tends to work against utility in those diseases, where an efficient reconfiguration of rigid networks needs to be achieved. This observation aligns with the documented value of NPs as sources of antibacterial and cytotoxic activity since it is much more difficult to develop resistance to a ligand targeting a more complex signaling node than for a simpler node. The centrality of a complex signaling node is likely to be associated with multiple pharmacological effects so the perturbation of a complex node may be more problematic for non-cytotoxic therapeutic approaches.

1.11. The physicochemical properties of NP drugs

The physicochemical properties of NP drugs as a function of the time period of discovery have been reviewed with the conclusion that lipophilicity and aromatic ring count were the most time invariant properties [41]. NPs as a class contain more macrocycles, a ring architecture of 12 or more atoms, than do collections of synthetic compounds designed for screening [42]. This observation aligns well with the observation that the biological information content for a cyclic compound is generally greater than for its acyclic counterpart [43]. Many types of biological activity require permeability across lipid membrane bilayers. To the extent that this is a powerful evolutionary selection pressure perhaps it might have influenced the NP biosynthesis toolkit. As a thought experiment, it is hard to imagine an NP biosynthesis toolkit evolving toward the excessively lipophilic and aqueous insoluble compound property profile exhibited by combinatorial synthetic drugs in the early years of HTS. I tend to think that the many cyclic structures, the multiple intramolecular hydrogen bonding possibilities and conformational mobility possibilities in NPs are structural features intended at least in part to increase membrane permeability possibilities.

2. Cyclosporine A—a NP prototype for a cyclic structure

Cyclosporine A (Fig. 1) is a NP prototype for a cyclic structure with, in theory, multiple intramolecular hydrogen bonding possibilities and conformational mobility possibilities. Based on chemical structure cyclosporine A is distinctly outside of Ro5 space, yet with formulation help it is orally bioavailable. The solvent dependent conformation of cyclosporine A has long been known [44]. In a series of elegant experimental and computational studies workers at Pfizer have shed further light on cyclosporine A's biophysical properties [45]. Based on NMR studies cyclosporine A adopts a markedly different intramolecular hydrogen bonding pattern and shape and polarity depending on whether it is found in aqueous polar or non-aqueous apolar medium.

Cyclosporine has 7 of the 11 peptide amide N–H′s in the cyclic peptide backbone replaced with *N*-methyls. In aqueous media the lipophilic *N*-methyls in the cyclic peptide are buried within the interior of the molecule and polar functionality is presented to the polar aqueous medium. In non-aqueous apolar medium all the lipophilic amide *N*-methyls project on the outside of the molecule into the apolar medium and the hydrophilic polar functionality is buried within the interior of the molecule.

Cyclosporine is literally a chemical chameleon changing its shape and polarity and its intramolecular hydrogen bonding pattern depending on whether it is in an aqueous or in a non-polar environment. This type of chameleon like behavior has been previously well documented for warfarin [46] and is likely fairly common in NPs although the number of experimentally well documented cases is low.

3. Ro5 and beyond

Drug discovery project decisions can lead to a non-NP Ro5 outlier. The orally active anti-cancer drug Navitoclax now in phase III studies provides an excellent example of the extensive expertise, resources, time and key decisions required for the rare discovery of a non-NP Ro5 outlier.

Protein–protein interactions (PPIs) pose a special challenge for the drug that conforms to the Ro5 parameter limits. For example, interfaces in PPI are usually flat and quite large (1000–2000 A^2). This compares to the deep clefts and cavities (300–500 A^2) that bind the smaller Ro5 compliant compound [47].

Analysis of the buried protein surface in a PPI suggests a minimum of about 500 A^2 is required for a stable dimeric protein association. As a consequence inhibitors of PPI have physicochemical properties outside the range of Ro5drugs and are likely to be more hydrophobic, more lipophilic and more likely to have a higher aromatic ring count than Ro5 compliant compounds. The bias against Ro5 non-compliant compounds in most HTS screening libraries is likely one factor in the failure of almost all HTS campaigns to discovery viable PPI inhibitor leads [48].

As ligand molecular size increases it is difficult to cover chemical space in screening libraries as large as a few million discrete compounds [49]. This limitation can be overcome to some extent by using DNA encoded libraries in which library sizes of several billion DNA tagged small molecules may be possible [50]. The PPI surface can be quite lipophilic and bland in the sense that the PPI depends on multiple small energetic interactions spread over a broad area rather than the crevice or hole "hot spot" found in a monomeric protein drug target. The PPI target poses a special problem in small molecule drug discovery since the property profile for a PPI drug, if it could be found, would likely be very non RO5-like. In fact, in the early years of HTS screening despite considerable screening effort, the success rate on finding anything useful in an HTS screen directed at a PPI target was close to zero. In recent times PPI targets have moved from undruggable to just druggable, albeit with very considerable effort, largely due to the advances in fragment screening [51] and from an increased understanding of the features of the protein–protein interface that lead to druggability [52].

3.1. Navitoclax—a non-NP Ro5 outlier

Navitoclax (Fig. 2) is an oncology drug discovered at Abbott that is in phase III clinical studies and that is directed at a PPI target that was

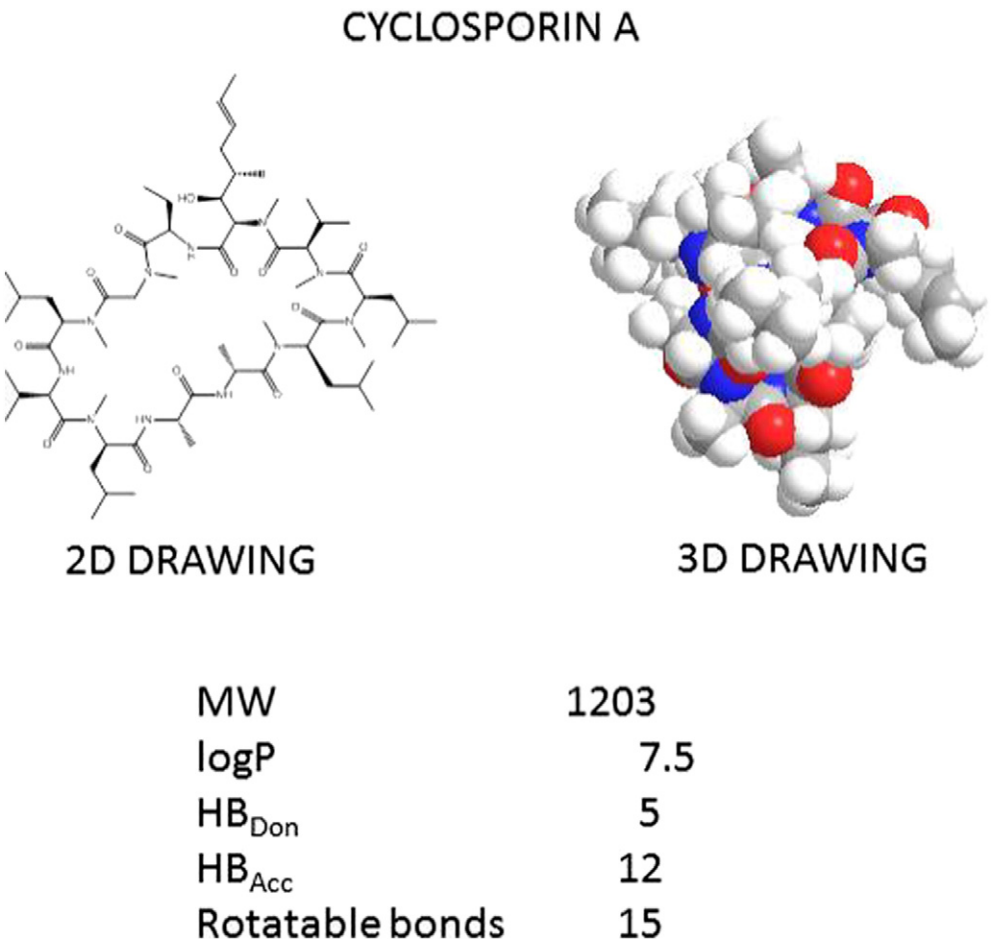

Fig. 1. Cyclosporin A. Representative Ro5 data were obtained from pubchem.ncbi.nlm.nih.gov. Molecular weight (MW), calculated partition coefficient between octanol and water (logP), hydrogen bond donors (HB_{Don}), hydrogen bond acceptors (HB_{Acc}) and number of rotatable bonds are shown.

discovered using fragment screening [53] and that based on chemical structure is very non Ro5 like. I would never in a million years have predicted any oral activity for Navitoclax based on its chemical structure. Nevertheless, with formulation help, the drug is orally bioavailable. Lymphatic uptake is responsible for much of the unexpected oral activity [54].

Navitoclax serves as a case study of a non NP RO5 outlier with outstanding documentation of the history of the steps and decision process taken by the Navitoclax drug discovery project team. The Navitoclax discovery story has been published in great detail in a Springer book chapter [53]. While a book chapter might in general be a bit less useful in allowing reader access than a scientific journal publication it has the advantage that the book format allows for a much richer discussion of what was really going on, what the discovery team was thinking and "what if" types of speculation concerning the management and project team's decision choices. The Navitoclax story not only illustrates that non NP oral drug activity is possible in a distinctly non Ro5 compound but also illustrates the very considerable skill, resources, and even luck that is required for an orally active non Ro5 compound.

My choice of a single take away message from the Navitoclax story is that timely and accurate pK/pD and oral absorption structure activity data enabled the medicinal chemistry use of the PK/PD relationship for AUC/EC_{50} [55] from a drug metabolism or pharmaceutical sciences screening group and that this data was absolutely critical to the discovery of this non NP and non Ro5 orally active drug.

It is by now clear that drugs can be discovered at various stages all along the way to clinical approval that lie well outside of Ro5 space. Numerous examples in the field of HCV therapy are compiled in a special thematic issue of the Journal of Medicinal Chemistry [56]. The discovery of the clinically approved Daclatasvir (MWT = 738.88) is a specially compelling example combining a phenotypic mechanistically unbiased screen instead of the more usual mechanistic screen; a very non-obvious structural progression from the original screening led to the final clinical candidate and issues of high molecular weight and log P [57].

4. Conclusions

Biologically active compounds cluster in small regions of the immensity of chemical space. Protein folds and pockets are limited in number. The physicochemical properties of drugs are controlled by the biophysics of formation of the cavities in proteins that are the ligand binding sites. Three of the four rule of five (Ro5) parameters, namely MWT, and hydrogen bond donors and acceptors fit well with how ligand binding sites are formed in proteins. Lipophilicity is a composite experimental solubility ratio parameter and does not directly relate to the biophysics of binding site formation. Natural products (NPs) provide most of the examples of Ro5 outliers. The special chemistry in NPs as well as their evolutionary selection allows the NP Ro5 outlier status. Some NPs like cyclosporine A are chemical chameleons, changing shape depending on the polarity of the medium. Discovery of a non NP Ro5 outlier like Navitoclax is very difficult but still is possible with a highly skilled, highly resourced team with excellent communication between medicinal chemists and in-vivo bioassay personnel. What is

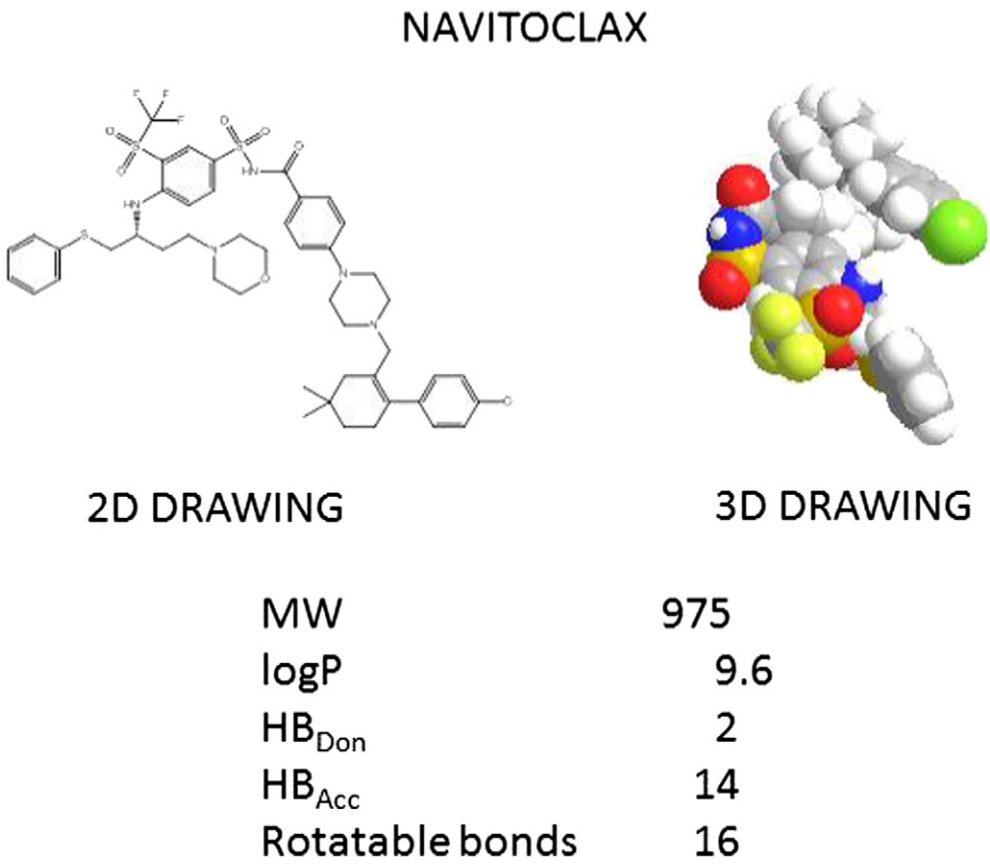

Fig. 2. Navitoclax. Representative Ro5 data were obtained from pubchem.ncbi.nlm.nih.gov. Molecular weight (MW), calculated partition coefficient between octanol and water (logP), hydrogen bond donors (HB_{Don}), hydrogen bond acceptors (HB_{Acc}) and number of rotatable bonds are shown.

unknown at this time is the cost-effectiveness of discovering non-natural product oral drugs well outside of Ro5 space. That such drugs can indeed be discovered is now well documented. That such drugs can be discovered efficiently is at this time quite unclear.

References

[1] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 23 (1997) 3–25.
[2] C. Lipinski, A. Hopkins, Navigating chemical space for biology and medicine, Nature 432 (2004) 855–861.
[3] N. Weskamp, E. Hüllermeier, G. Klebe, Merging chemical and biological space: structural mapping of enzyme binding pocket space, Proteins: Struct., Funct., Bioinf. 76 (2009) 317–330.
[4] S. Barelier, T. Sterling, M.J. O'Meara, B.K. Shoichet, The recognition of identical ligands by unrelated proteins, ACS Chem. Biol. 10 (2015) 2772–2784.
[5] J. Hert, M.J. Keiser, J.J. Irwin, T.I. Oprea, B.K. Shoichet, Quantifying the relationships among drug classes, J. Chem. Inf. Model. 48 (2008) 755–765.
[6] D.D. Axe, Estimating the prevalence of protein sequences adopting functional enzyme folds, J. Mol. Biol. 341 (2004) 1295–1315.
[7] C.M. Dobson, Chemical space and biology, Nature 432 (2004) 824–828.
[8] D.J. Triggle, The chemist as astronaut: searching for biologically useful space in the chemical universe, Biochem. Pharmacol. 78 (2009) 217–223.
[9] Anonymous, Growth of unique folds per year as defined by SCOP (v1.75), Protein Data Bank, http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=fold-scop
[10] T. Sikosek, H.S. Chan, Biophysics of protein evolution and evolutionary protein biophysics, J. R. Soc. Interface 11 (2014) 20140419.
[11] M. Weisel, E. Proschak, J.M. Kriegl, G. Schneider, Form follows function: shape analysis of protein cavities for receptor-based drug design, Proteomics 9 (2009) 451–459.
[12] R. Abagyan, C.G. Ricci, in: Gisbert Schneider (Ed.), The Human Pocketome, De Novo Molecular Design, first ed.Wiley-VCH 2014, pp. 73–96 (Chapter 3).
[13] E. Perola, P.S. Charifson, Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding, J. Med. Chem. 47 (2004) 2499–2510.
[14] M. Bartolowits, V. Davisson, Considerations of protein subpockets in fragment-based drug design, Chem. Biol. Drug Des. 87 (2016) 5–20.
[15] Binding of small molecules to cavity forming mutants of a de novo designed protein, A. Das, Y. Wei, I. Pelczer, M.H. Hecht, Protein Sci. 20 (2011) 702–711.
[16] S.C. Patel, L.H. Bradley, S.P. Jinadasa, M.H. Hecht, Cofactor binding and enzymatic activity in an unevolved superfamily of de novo designed 4-helix bundle proteins, Protein Sci. 18 (2009) 1388–1400.
[17] J. Skolnick, M. Gao, Interplay of physics and evolution in the likely origin of protein biochemical function, PNAS 110 (2013) 9344–9349.
[18] R.J. Zauhar, E. Gianti, W.J. Welsh, Fragment-based shape signatures: a new tool for virtual screening and drug discovery, J. Comput. Aided Mol. Des. 27 (2013) 1009–1036.
[19] R.D. Cramer, R.J. Jilek, S. Guessregen, S.J. Clark, B. Wendt, R.D. Clark, "Lead hopping". Validation of topomer similarity as a superior predictor of similar biological activities, J. Med. Chem. 47 (2004) 6777–6791.
[20] A. Gershenson, L.M. Gierasch, A. Pastore, S.E. Radford, Energy landscapes of functional proteins are inherently risky, Nat. Chem. Biol. 10 (2014) 884–891.
[21] A.W. Fitzpatrick, T.P. Knowles, C.A. Waudby, M. Vendruscolo, C.M. Dobson, Inversion of the balance between hydrophobic and hydrogen bonding interactions in protein folding and aggregation, PLoS Comput. Biol. 7 (2011) 1002169.
[22] S. Pérot, O. Sperandio, M.A. Miteva, A. Camproux, B.O. Villoutreix, Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery, Drug Discov. Today 15 (2010) 656–667.
[23] G. Wei, W. Xi, R. Nussinov, B. Ma, Protein ensembles: how does nature harness thermodynamic fluctuations for life? The diverse functional roles of conformational ensembles in the cell, Chem. Rev. (2010), http://dx.doi.org/10.1021/acs.chemrev.5b00562.
[24] M. Andrec, D.A. Snyder, Z. Zhou, J. Young, G.T. Montelione, R.M. Levy, A large data set comparison of protein structures determined by crystallography and NMR: statistical test for structural differences and the effect of crystal packing, Proteins: Struct., Funct., Bioinf. 69 (2007) 449–465.
[25] G.G. Ferenczy, G.M. Keseru, On the enthalpic preference of fragment binding, Med. Chem. Commun. 7 (2016) 332–337.
[26] Y. Kawasaki, E. Chufan, V. LaFont, K. Hidaka, Y. Kiso, L. Mario Amzel, E. Freire, How much binding affinity can be gained by filling a cavity? Chem. Biol. Drug Des. 75 (2010) 143–151.
[27] E. Freire, in: Waksman (Ed.), A Thermodynamic Guide to Affinity Optimization of Drug Candidates, Proteomics and Protein–Protein Interactions: Biology, Chemistry, Bioinformatics, and Drug Design, Springer, New York 2005, pp. 291–307 (Chapter 13).
[28] J. Böstrom, D. Brown, Stuck in a rut with old chemistry, Drug Discov. Today (2016), http://dx.doi.org/10.1016/j.drugdis.2016.02.017.
[29] P.M. Dewick, Secondary Metabolism: the Building Blocks and Construction Mechanisms, Medicinal Natural Products: a Biosynthetic Approach, John Wiley & Sons,

Ltd, 2001 7–34 (ISBNs: 0471496405 (hardback); 0471496413 (paperback); 0470846275 (electronic)).
[30] C.A. Lipinski, N.K. Litterman, C. Southan, A.J. Williams, A.M. Clark, S. Ekins, Parallel worlds of public and commercial bioactive chemistry data, J. Med. Chem. 58 (2015) 2068–2076.
[31] G. Müller, H. Giera, Protein secondary structure templates derived from bioactive natural products—combinatorial chemistry meets structure-based design, J. Comput. Aided Mol. Des. 12 (1998) 1–6.
[32] S.F. Hahk-Soo Kang, Brady, mining soil metagenomes to better understand the evolution of natural product structural diversity: pentangular polyphenols as a case study, J. Am. Chem. Soc. 136 (2014) 18111–18119.
[33] B.M. McArdle, M.R. Campitelli, R.J. Quinn, A common protein fold topology shared by flavonoid biosynthetic enzymes and therapeutic targets, J. Nat. Prod. 69 (2006) 14–17.
[34] N. Sturm, R.J. Quinn, E. Kellenberger, Similarity between flavonoid biosynthetic enzymes and flavonoid protein targets captured by three-dimensional computing approach, Planta Med. 81 (2015) 467–473.
[35] M.A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casaulta, A. Odermatt, P. Ertl, H. Waldmann, Charting biologically relevant chemical space: a structural classification of natural products (SCONP), Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 17272–17277.
[36] K. Grabowski, G. Schneider, Properties and architecture of drugs and natural products revisited, Curr. Chem. Biol. 1 (2007) 115-12.
[37] B.C. Doak, J. Zheng, D. Dobritzsch, J. Kihlberg, How beyond rule of 5 drugs and clinical candidates bind to their targets, J. Med. Chem. 59 (2016) 2312–2327.
[38] J. Bisson, J.B. McAlpine, J.B. Friesen, S. Chen, J. Graham, G.F. Pauli, Can invalid bioactives undermine natural product-based drug discovery? J. Med. Chem. (2015 Oct 27) (epub ahead of print).
[39] G.V. Dancˇi'k, K.P. Seiler, D.W. Young, S.L. Schreiber, P.A. Clemons, Distinct biological network properties between the targets of natural products and disease, J. Am. Chem. Soc. 132 (2010) 9259–9261.
[40] P. Csermely, T. Korcsmáros, H.J. Kiss, G. London, R. Nussinov, Structure and dynamics of molecular networks: a novel paradigm of drug discovery a comprehensive review, Pharmacol. Ther. 138 (2013) 333–408.
[41] D. Camp, A. Garavelas, M. Campitelli, Analysis of physicochemical properties for drugs of natural origin, J. Nat. Prod. 78 (2015) 1370–1382.
[42] E.M. Driggers, S.P. Hale, J. Lee, N.K. Terrett, The exploration of macrocycles for drug discovery—an underexploited structural class, Nat. Rev. Drug Discov. 7 (2008) 608–624.
[43] Kim Young-kwon, E.F. Dean III, M.A. Arai, T. Arai, N. Patterson, J.O. Lamenzo, P.A. Clemons, S.L. Schreiber, Relationship of Stereochemical and skeletal diversity of small molecules to cellular measurement space, J. Am. Chem. Soc. 126 (2004) 14740–14745.
[44] N. El Tayar, A.E. Mark, P. Vallat, R.M. Brunne, B. Testa, W.F. van Gunsteren, Solvent-dependent conformation and hydrogen-bonding capacity of cyclosporin a: evidence from partition coefficients and molecular dynamics simulations, J. Med. Chem. 36 (1993) 3757–3764.
[45] A. Alex, D.S. Millan, M. Perez, F. Wakenhut, G.A. Whitlock, Intramolecular hydrogen bonding to improve membrane permeability and absorption in beyond rule of five chemical space, Med. Chem. Commun. 2 (2011) 669–674.
[46] I.A. Nicholls, B.C.G. Karlsson, A.M. Rosengren, H. Henschel, Warfarin: an environment-dependent switchable molecular probe, J. Mol. Recognit. 23 (2010) 604–608.
[47] M.R. Arkin, Y. Tang, J.A. Wells, Small-molecule inhibitors of protein–protein interactions: progressing toward the reality, Chem. Biol. 21 (2014) 1102–1114.
[48] M.A. Kuenemann, L.M. Bourbon, C.M. Labbé, B.O. Villoutreix, O. Sperandio, Which 3D characteristics make an efficient inhibitor of protein–protein interactions? J. Chem. Inf. Model. 54 (2014) 3067–3079.
[49] M.M. Hann, Molecular obesity, potency and other addictions in drug discovery, Med. Chem. Commun. 2 (2011) 349–355.
[50] R.M. Franzini, D. Neri, J. Scheuermann, DNA-encoded chemical libraries: advancing beyond conventional small-molecule libraries, Acc. Chem. Res. 47 (2014) 1247–1255.
[51] B.O. Villoutreix, C. Labbé, D. Lagorce, G. Laconde, O. Sperandio, A leap into the chemical space of protein–protein interaction inhibitors, Curr. Pharm. Des. 30 (2012) 4648.
[52] D.K. Johnson, J. Karanicolas, Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface, PLoS Comput. Biol. 9 (3) (2013) e1002951, http://dx.doi.org/10.1371/journal.pcbi.1002951.
[53] M.D. Wendt, The discovery of navitoclax, a Bcl-2 family inhibitor, protein–protein interactions, Topics Med. Chem. 8 (2012) 231–258.
[54] E.F. Choo, J. Boggs, C. Zhu, J.W. Lubach, N.D. Catron, G. Jenkins, A.J. Souers, R. Voorman, The role of lymphatic transport on the systemic bioavailability of the Bcl-2 protein family inhibitors navitoclax (ABT-263) and ABT-199, Drug Metab. Dispos. 42 (2014) 207–212.
[55] Park Cheol-Min, M. Bruncko, J. Adickes, J. Bauch, H. Ding, A. Kunzer, K.C. Marsh, P. Nimmer, A.R. Shoemaker, X. Song, S.K. Tahir, C. Tse, X. Wang, M.D. Wendt, X. Yang, H. Zhang, S.W. Fesik, S.H. Rosenberg, S.W. Elmore, Discovery of an orally bioavailable small molecule inhibitor of prosurvival B-cell lymphoma 2 proteins, J. Med. Chem. 51 (2008) 6902–6915.
[56] N.A. Meanwell, W.J. Watkins, Introduction to hepatitis C virus (HCV) therapies special thematic issue, J. Med. Chem. 57 (2014) 1625–1626.
[57] M. Belema, N.A. Meanwell, Discovery of daclatasvir, a pan-genotypic hepatitis C virus NS5A replication complex inhibitor with potent clinical effect, J. Med. Chem. 57 (2014) 5057–5071.