**Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs beyond the Rule of 5**

# Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs Beyond the Rule of 5

Matteo Rossi Sebastiano,[1,a] Bradley C. Doak,[2,a] Maria Backlund,[3] Vasanthanathan Poongavanam,[1] Björn Over,[4] Giuseppe Ermondi,[5] Giulia Caron,[5] Pär Matsson[6,*] and Jan Kihlberg[1,*]

[1]Department of Chemistry - BMC, Uppsala University, Box 576, SE-751 23 Uppsala, Sweden

[2]Department of Medicinal Chemistry, MIPS, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia

[3]Uppsala University Drug Optimization and Pharmaceutical Profiling Platform (UDOPP), a node at the Chemical Biology Consortium Sweden, Science for Life Laboratory, Department of Pharmacy, BMC, Uppsala University, Box 580, SE-751 23 Uppsala, Sweden

[4]Cardiovascular and Metabolic Diseases, Innovative Medicines and Early Development Biotech Unit, AstraZeneca R&D Gothenburg, SE-431 83 Mölndal, Sweden

[5]Department of Molecular Biotechnology and Health Sciences, University of Torino, Quarello 15, 10135, Torino, Italy

[6]Department of Pharmacy, BMC, Uppsala University, Box 580, SE-751 23 Uppsala, Sweden

[a]Equally contributing authors

*Corresponding authors: par.matsson@farmaci.uu.se, jan.kihlberg@kemi.uu.se

**Keywords:** Beyond rule of 5, macrocycle, polar surface area, conformational flexibility, permeability, solubility, molecular chameleon

**ABSTRACT:** Conformational flexibility has been proposed to significantly affect drug properties outside rule-of-5 (Ro5) chemical space. Here, we investigated the influence of dynamically exposed polarity on cell permeability and aqueous solubility for a structurally diverse set of drugs and clinical candidates far beyond the Ro5, all of which populated multiple distinct conformations as revealed by X-ray crystallography. Efflux-inhibited (passive) Caco-2 cell permeability correlated strongly with the compounds' minimum solvent-accessible 3D polar surface areas (PSA), while aqueous solubility depended less on the specific 3D conformation. Inspection of the crystal structures highlighted flexibly linked aromatic side chains and dynamically forming intramolecular hydrogen bonds as particularly effective in providing "chameleonic" properties that allow compounds to display both high cell permeability and aqueous solubility. These structural features, in combination with permeability predictions based on the correlation to solvent-accessible 3D PSA, should inspire drug design in the challenging chemical space far beyond the Ro5.

## INTRODUCTION

Improved selection of novel disease-associated drug targets has been highlighted as the most important factor for success in drug discovery.[1] Unfortunately, half of all targets assumed to be involved in human disease have been classified as "difficult to drug"[2, 3] with traditional small molecules, i.e. ligands that reside in the chemical space defined by Lipinski´s rule of 5 (Ro5).[4, 5] However, recent investigations have revealed that macrocycles,[6-10] and other compounds outside

Ro5 chemical space,[11, 12] provide improved opportunities for modulation of difficult to drug targets. In particular, compounds residing in what has been termed beyond-rule-of-5 (bRo5) space[12, 13] are better suited for modulation of targets that have large, flat and groove-shaped binding sites, for instance protein-protein interactions. Understanding how to design and optimize orally administered compounds in bRo5 space is therefore essential for future success in drug discovery, especially for intracellular targets that are not accessible to biologics.

Recently, three datasets have been analyzed to provide insight into the outer limits of chemical space in which pharmacokinetic risks may be managed and novel cell-permeable and orally bioavailable drugs have a reasonable chance of being discovered.[14-16] The datasets consist of *i*) approved drugs and clinical candidates in bRo5 space,[14] *ii*) carefully designed libraries of cyclic peptides[15] and *iii*) compounds from preclinical drug discovery projects.[16] High-level analyses of these datasets have revealed that MW may be increased up to approximately 1,000 Da, topological polar surface area (TPSA) to 250 Å$^2$, the number of rotatable bonds to 20 and the number of hydrogen bond acceptors (HBA) to 15. However, lipophilicity should be controlled in drug-like space (e.g. cLogP between 3 and 6), in particular at high MW,[15, 16] and few orally available drugs have more than six hydrogen bond donors (HBD).[14, 16] More detailed analyses of the same datasets,[14-16] additional sets of macrocyclic peptides,[17-25] and de novo designed macrocycles inspired by natural products[26, 27] have also begun to reveal the molecular properties, structural features and functional groups that allow cell permeability in this non-traditional drug space. These studies have highlighted reduction of polarity by N-alkylation of solvent-exposed amide bonds, shielding of polar groups by bulky side chains and induction of intramolecular hydrogen bonds (IMHB) as effective tactics that may be used to increase cell permeability.

However, these tactics often suffer from the drawback of decreasing solubility in aqueous media.[28]

Case studies of cyclosporin A,[29] a cyclic peptide model system,[22] and a set of stereoisomeric de novo designed macrocycles[26] have revealed that bRo5 compounds that dynamically expose or shield polarity can combine high permeability and aqueous solubility, in contrast to analogues that are rigid or for which the molecular properties do not vary among conformations.[27] Conformational flexibility may thus provide "chameleonic", environment-dependent properties to compounds, for instance by enabling them to transiently form IMHBs and present a less polar surface when crossing the lipophilic cell membrane, while a more polar surface with exposed hydrogen bonding functionalities is displayed in aqueous environments.[29-36] Recent analyses of crystal structures of drugs in bRo5 space have revealed significant conformational differences that result in large variations in the 3D polar surface area (PSA), suggesting that conformational flexibility may be of general importance for compound properties in this space.[12, 28, 30] It was proposed that adequate solubility in bRo5 space requires compounds to have a PSA of $\geq 0.2 \times$ MW,[30] while early studies in the field have revealed that PSA should be $\leq 140$ Å$^2$ for satisfactory passive cell permeability.[37, 38] These two guidelines indicate that compounds with MW >700 Da must be chameleonic, i.e. be able to adapt their PSA to the environment, in order to satisfy both criteria.[28, 30] This appealing hypothesis has, however, not been experimentally tested beyond the few specific examples listed above.[22, 26, 29]

Here, our aim was to investigate to what extent conformation-dependent, dynamically exposed polarity provides chameleonic physicochemical properties to orally available drugs in bRo5 space, and what substructural features stand out as important for such chameleonic behavior. First, we determined how 2D PSA (TPSA; calculated directly from the connectivity of a

compound) and 3D PSA (derived from 3D conformers) varied with increasing MW for drugs listed in the DrugBank database, so as to get an overview of how polarity is exposed or hidden in different regions of chemical space. Then we selected 24 compounds representing the major chemical classes of orally available drugs and clinical candidates,[14] covering both macrocyclic and non-macrocyclic structures of natural product-derived and *de novo*-designed origins. The compounds showed varying degrees of conformational flexibility as revealed by multiple X-ray crystal structures. We identified a training set of eleven compounds and measured cell permeability, solubility, lipophilicity and $pK_a$ under consistent experimental conditions so that data were unaffected by inter-laboratory variation. Quantitative models of how exposed polarity of experimentally confirmed conformations correlated to cell permeability and solubility were derived and validated using literature data. Finally, inspection of the crystal structures for all 24 drugs and clinical candidates provided an assessment of the nature and prevalence of dynamically formed intramolecular interactions that result in variations of 3D PSA.

## RESULTS AND DISCUSSION

**Conformational Flexibility and Burying of Polar Surface Area in bRo5 Space.** The number of rotatable bonds (NRotB) provides an indication of a compound's conformational flexibility. Analysis of approved orally administered drugs reveals, as may have been expected, that NRotB increases with increasing MW (Figure 1A). Oral drugs in bRo5 space have a significantly higher mean NRotB than oral drugs with MW 100-500 Da ($p < 0.0001$). In addition, all oral drugs in bRo5 space have $\geq$5 NRotB, highlighting that oral drugs in this chemical space show some degree of conformational flexibility, which could provide them with chameleonic physicochemical properties.

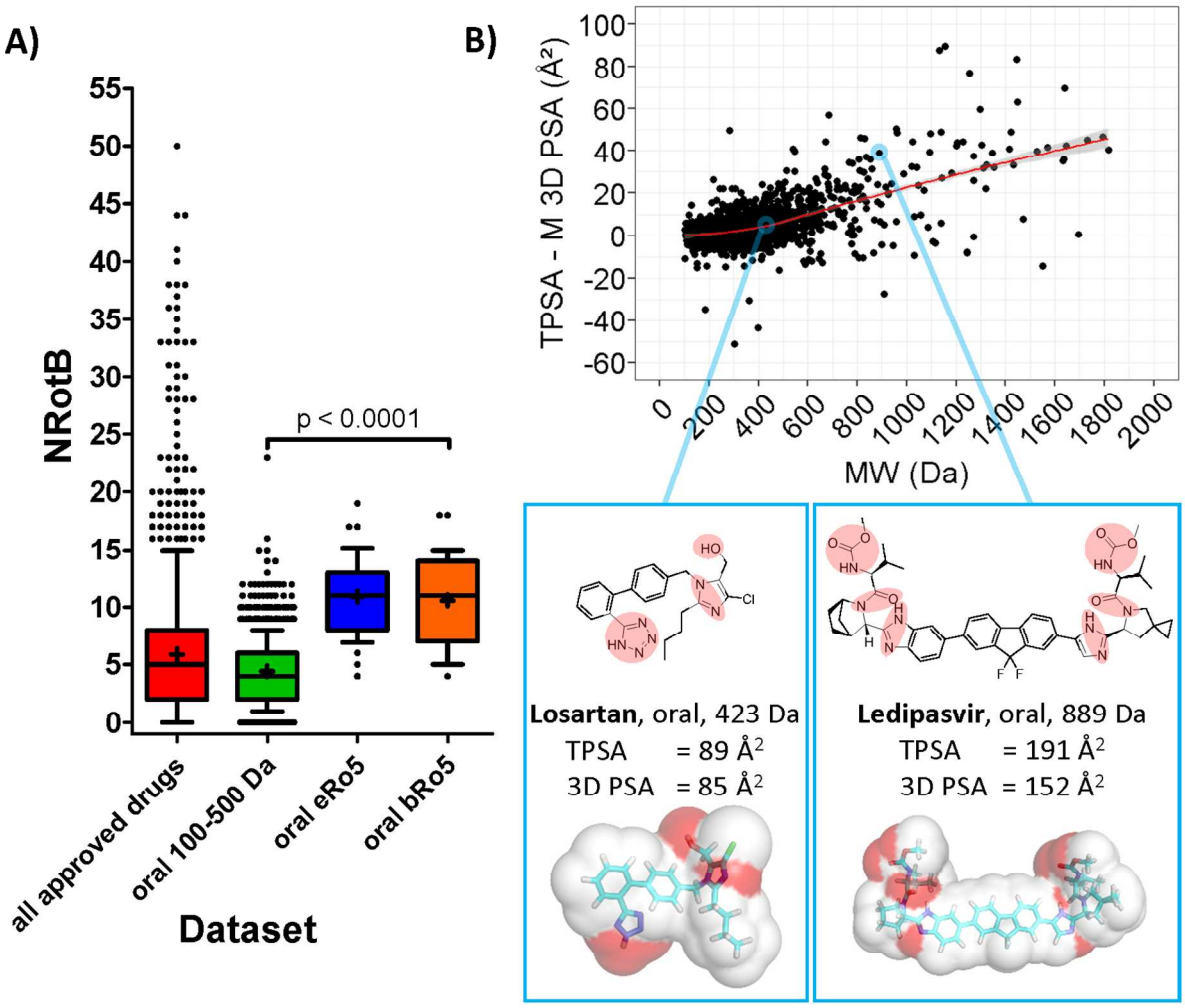



**Figure 1.** A) Distribution of number of rotatable bonds (NRotB) for all approved drugs in DrugBank (mean 5.9, stdev 5.5, n=1814) followed by orally administered, approved drugs having MWs of 100-500 Da from Drugbank (mean 4.4 stdev 3.0, n=788), oral drugs in eRo5 space[14] (500-700 Da; mean 10.9, stdev 3.3, n=38) and oral drugs in bRo5 space[14] (>700 Da; mean 10.6 stdev 3.9, n=34, $p < 0.0001$ for t-test comparison to approved oral 100-500 Da drugs). Whiskers and boxes show the 10$^{th}$, 25$^{th}$, 50$^{th}$, 75$^{th}$ and 90$^{th}$ percentiles, means are shown as a crosses and outliers as black dots. B) Difference between 2D TPSA and molecular 3D PSA (M 3D PSA) for calculated low energy conformations of all approved drugs in the DrugBank

database with molecular weight (MW) 100-2000 Da, plotted versus MW. The data is fitted to a local linear regression function (LOESS, red line), showing how the difference in polar surface area changes with increasing MW. The regression standard error is shown as gray shading. Structures and molecular properties are included for losartan and ledipasvir, as examples of drugs in Ro5 and bRo5 space, respectively.

Polar surface area is an established predictor of cell permeability[37, 38] and, as discussed above, a conformation-dependent variation in PSA may be essential for drug-like properties in bRo5 space.[12, 28, 30, 39] To evaluate if PSA increasingly depends on conformational preferences with increasing MW, we calculated TPSA for all approved drugs having MWs of 100-2000 Da in the DrugBank database[40] and compared it to a 3D molecular PSA calculated for the predicted low energy conformations using the same atom and surface definition as for TPSA (Figure 1).[41] For typical small molecule drugs (MW <500 Da; exemplified by losartan in Figure 1), the PSA predicted from 2D fragment contributions (TPSA) expectedly corresponds well to molecular PSA calculated from 3D structures (standard deviation = 5.8 $Å^2$ for drugs 100-500 Da). Hence, and also since it is rapidly calculated for large compound libraries, TPSA has found widespread use in Ro5 drug discovery. However, at MWs above 500 Da the difference between TPSA and molecular 3D PSA begins to increase linearly, indicating a conformation-dependent variability in PSA that is not captured by TPSA. Thus, in general, polarity is buried to an increasing degree as molecular size, structural complexity and often also flexibility increases farther into bRo5 space,[42] as exemplified by ledipasvir (Figure 1). This observation provides initial support for the hypothesis that dynamic exposure of polarity may be of general importance for physicochemical properties in bRo5 space.

**Compound and Structural Data Selection.** Because of the difficulty of accurately predicting the conformational landscape for compounds outside the Ro5 by computational methods,[43, 44] we assessed compound flexibility from available X-ray crystal structures. Arguably, these may not exhaustively cover the respective conformational landscapes because of crystal packing effects and limitations in the availability of structural data.[45, 46] However, the approach has the clear benefit of being directly based on experimentally verified conformations.[12, 25, 34] Out of the 11 major chemical classes of orally available drugs and clinical candidates in bRo5 space[14] we therefore selected a ***crystal structure set*** consisting of 24 compounds, originating from seven of the 11 major classes plus one additional cyclic peptide and one natural product (Table 1). Among these 24 compounds we also selected a ***training set*** and a ***test set*** consisting of 11 and eight compounds, respectively. Cell permeability, solubility and LogD were determined for the drugs and clinical candidates in the training set, while cell permeability was retrieved from the literature for the compounds in the external test set.

Compounds were included in the crystal structure dataset if at least two high resolution crystal structures were available in the Protein Data Bank (PDB, <3.5 Å resolution) or the Cambridge Structural Database (CSD) that represented different conformations (heavy atom RMSD >0.75 Å between conformers). The crystal structures from the PDB were inspected to ensure that all conformations had well defined electron densities, high occupancies, temperature factors similar to target side chains adjacent to the ligand in the binding site and unstrained geometries. Overall, the 24 compounds in the crystal structure dataset included six orally administered erythronolides, six HIV-1 protease inhibitors, three HCV NS3/4A protease inhibitors, four rifamycins, two cyclic peptides and three additional natural products from different bRo5 classes (Table 1; Table

S1). The 24 compounds had MWs that were evenly distributed between 500 and 900 Da, except for the two cyclic peptides which had significantly higher MWs. TPSA increased proportionally with MW from 120 to 360 Å$^2$,[14, 30] while cLogP remained centered at 3-5 (Figures 2, and Figure S1). These relationships for TPSA and cLogP are also characteristic for the comprehensive list of 85 orally administered drugs and clinical candidates identified in bRo5 space;[14] thus the selected compounds provide a representative and structurally diverse coverage of this property space. The number of crystal structures retrieved for each of the compounds ranged from two to 51, all of which were retained for subsequent analyses. After clustering similar conformations (RMSD <0.75 Å) the numbers of distinct conformations ranged from two to five for the different compounds. The effective flexibility assessed, i.e. the sampled conformational space, varied from low (heavy atom RMSD = 0.82 Å between the most different conformers) for some compounds to significantly larger (6.13 Å) for others. The crystal structure set, as well as the training and test sets contained within it, was therefore selected to display a range of conformational flexibilities, as revealed by the experimentally determined X-ray structures.

**Table 1.** Structural and conformational features of the compounds included in the crystal structure dataset,[a] the training set[b] and the external test set[c]

| Class | Name[a,b,c] | macrocycle | n° structures | n° conformers[d] | Max RMSD[e] (Å) | MW (Da) | TPSA[f] (Å$^2$) |
|---|---|---|---|---|---|---|---|
| Erythronolides | **Azithromycin** | Y | 7 | 3 | 3.14 | 749 | 180 |
| | **Clarithromycin** | Y | 14 | 3 | 3.13 | 748 | 183 |
| | Dirithromycin | Y | 4 | 2 | 0.82 | 835 | 194 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Erythromycin** | Y | 10 | 4 | 1.94 | 734 | 194 |
| | **Roxithromycin** | Y | 3 | 2 | 3.96 | 837 | 217 |
| | **Telithromycin** | Y | 5 | 2 | 6.13 | 812 | 172 |
| HIV-1 Inhibitor | **Atazanavir** | | 12 | 4 | 3.30 | 705 | 171 |
| | Darunavir | | 51 | 5 | 3.78 | 548 | 140 |
| | *Indinavir* | | 16 | 4 | 2.00 | 614 | 119 |
| | Lopinavir | | 10 | 4 | 1.72 | 629 | 120 |
| | **Ritonavir** | | 17 | 5 | 3.96 | 721 | 146 |
| | **Saquinavir** | | 29 | 3 | 3.90 | 671 | 168 |
| NS3/4A Inhibitor | **Asunaprevir** | | 4 | 2 | 3.02 | 748 | 194 |
| | *Faldaprevir* | | 2 | 2 | 4.54 | 870 | 201 |
| | **Telaprevir** | | 5 | 3 | 1.45 | 680 | 180 |
| Rifamycin | Rifabutin | Y | 2 | 2 | 1.22 | 847 | 210 |
| | **Rifampicin** | Y | 8 | 3 | 1.56 | 823 | 220 |
| | Rifapentin | Y | 2 | 2 | 0.98 | 877 | 220 |
| | *Rifaximin* | Y | 7 | 2 | 1.64 | 786 | 201 |
| Cyclic peptides | *Actinomycin D* | Y | 12 | 2 | 3.92 | 1255 | 356 |
| | *Cyclosporin A* | Y | 18 | 9 | 5.97 | 1203 | 279 |
| Other natural products | *Tacrolimus* | Y | 6 | 5 | 3.72 | 804 | 178 |
| | *Paclitaxel* | Y | 3 | 2 | 3.51 | 854 | 221 |
| | *Vinblastine* | | 4 | 2 | 0.92 | 811 | 157 |

[a]All compounds were included in the crystal structure dataset. [b]The compounds in bold were included in the training set and their permeability across Caco-2 cell monolayers, solubility, logD and $pK_a$ were determined as described in the Experimental Section. [c]The compounds in italics were used as an external test set with efflux-inhibited cell permeabilities across Caco-2 cell monolayers compiled from the literature.[47-53] [d]The number of distinct conformers was determined by clustering the available crystal structures, using a threshold of 0.75 Å for the root

mean squared deviation (RMSD) of the heavy atom alignment. [e]Maximum RMSD of the heavy atom alignment of crystal structures representing the two most diverse conformations. [f]TPSA was calculated using the method reported by Ertl, et al.[41] as implemented in Instant J Chem v6.2.0.953. Crystal structure database identifiers for all evaluated structures are available in Table S1.

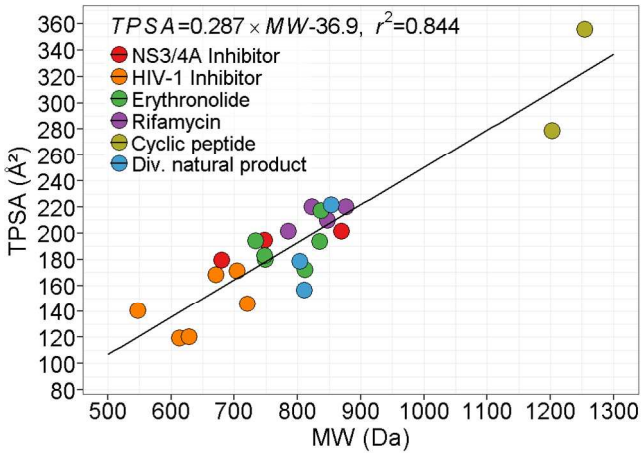



**Figure 2.** Topological polar surface area (TPSA, Å$^2$) versus molecular weight (MW, Da) for the selected erythronolides, rifamycins, HIV-1, HCV NS3/4A protease inhibitors, cyclic peptides and other natural products. TPSA was calculated using the method reported by Ertl, et al.[41] as implemented in Instant J Chem v6.2.0.953.

The 11 compounds in the training set were obtained from the erythronolide and rifamycin classes of antibacterials, and the *de novo* designed antiviral HIV-1 protease and HCV NS3/4A protease inhibitors (in boldface in Table 1). This selection was made as multiple crystal structures are available for several members of these four structural classes. In addition, most members of these

classes were commercially available, allowing us to generate in-house solubility, lipophilicity and cell permeability data for them under consistent conditions. The selection led to that macrocyclic as well as non-macrocyclic compounds, originating from natural products as well as from structure-based design, were included in the training set. The 8 compounds in the external test set (in italics in Table 1) included compounds from three of the four classes in the training set, compounds from an additional three of the major bRo5 classes (the cyclosporins, ascomycins and taxanes) as well as one additional cyclic peptide (actinomycin D) and one additional natural product (vinblastine).[47-53] Compounds were included in the test set if their efflux-inhibited cell permeability had been reported in the literature, and if multiple crystal structures were available for each of them. All but three of the compounds in the training and tests sets had MWs in the range where achieving both good aqueous solubility and cell permeability has been proposed to be contingent on conformational flexibility, i.e. in the difficult bRo5 space at 700 Da or above. The 16 compounds with MW >700 Da constitute 19% of all orally administered drugs and clinical candidates found in bRo5 space in our recent analysis.[14]

**Cell Permeability.** Permeability across intestinal epithelial cell monolayers was measured in the Caco-2 cell model for the 11 training set compounds. Monolayer permeabilities ($P_{app}$) were determined at pH 7.4 in the apical-to-basolateral (AB) and basolateral-to-apical (BA) directions, and efflux ratios ($P_{app}$ BA / $P_{app}$ AB) were used to assess the impact of carrier-mediated transport processes (Table 2). In addition, passive-diffusive cell monolayer permeabilities ($P_{app}$ AB +inh and $P_{app}$ BA +inh) were estimated by including an inhibitor cocktail of quinidine, sulfasalazine and benzobromarone that targets the three major efflux transporters in the intestinal epithelium: P-glycoprotein, (MDR1/P-gp; ABCB1), breast cancer resistance protein (BCRP; ABCG2) and

multidrug-resistance associated protein 2 (MRP2, ABCC2). Cell permeabilities determined in the absence of inhibitors ($P_{app}$ AB) ranged from low (<1 × $10^{-6}$ cm/s) for saquinavir, telaprevir and three of the erythronolides to medium (10.9 × $10^{-6}$ cm/s) for atazanavir. All compounds were significantly affected by apical efflux transporters, with efflux ratios (ER) ranging from 11 for azithromycin to >500 for saquinavir; this considerable transporter-mediated efflux was consistent with previous analyses of cellular permeability in bRo5 space.[11, 28] Accordingly, apical-to-basolateral cell permeabilities determined in the presence of the inhibitor cocktail were significantly higher than the corresponding uninhibited permeabilities for all compounds. Residual efflux was observed, to different extents, also in the presence of the inhibitor cocktail (Table 2). This was particularly pronounced for saquinavir (residual ER 27.8); thus, the permeability estimate is not representative of the passive transcellular permeability and saquinavir was, consequently, excluded from the modeling of passive cell permeability.

**Table 2.** Experimental Caco-2 permeability data for the bRo5 drugs and clinical candidates included in the training set.

| Class | Name | $P_{app}$ AB[a] (×$10^{-6}$ cm/s) | SEM | ER[b] | $P_{app}$ AB +inh[a,c] (×$10^{-6}$ cm/s) | SEM | ER +inh[b,c] |
|---|---|---|---|---|---|---|---|
| Erythronolides | Azithromycin | 1.8 | 0.4 | 11 | 2.8 | 0.1 | 7.9 |
| Erythronolides | Clarithromycin | 2.7 | 0.3 | 28 | 28.2 | 1.3 | 1.2 |
| Erythronolides | Erythromycin | 0.58 | 0.03 | 33 | 1.5 | 0.1 | 1.5 |
| Erythronolides | Roxythromycin | 0.90 | 0.04 | 57 | 11.9 | 1.6 | 1.0 |
| Erythronolides | Telithromycin | 0.20 | 0.02 | 203 | 4.3 | 0.1 | 3.9 |
| HIV-1 inh. | Atazanavir | 10.9 | 0.2 | 40 | 71.8 | 2.1 | 2.3 |
| HIV-1 inh. | Ritonavir | 8.7 | 0.3 | 142 | 232 | 40 | 2.1 |
| HIV-1 inh. | *Saquinavir*[d] | 0.52 | 0.1 | 585 | 5.0 | 0.2 | 28 |
| NS3/4A inh. | Asunaprevir | 3.4 | 0.4 | 272 | 48.0 | 5.7 | 4.1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| NS3/4A inh. | Telaprevir | 0.41 | 0.1 | 199 | 7.1 | 0.4 | 4.0 |
| Rifamycin | Rifampicin | 1.6 | 0.02 | 19 | 1.0 | 0.1 | 14.4 |

[a]$P_{app}$ AB: permeability in the apical-to-basolateral (AB) direction across Caco-2 cell monolayers. [b]ER: efflux ratio ($P_{app}$ BA / $P_{app}$ AB). [c]Determined in the presence of a cocktail of inhibitors (quinidine, sulfasalazine and benzobromarone, noted by +inh) of efflux transporters. [d]Excluded from modeling of permeability. All measurements were performed in triplicate, except for saquinavir and ritonavir (n=6, repeated with consistent results because of excessive mass balance in the uninhibited experiment). Measurements in both transport directions were performed both in the absence and presence of a cocktail of efflux transporter inhibitors.

**Solubility, Lipophilicity and $pK_a$.** Solubility was determined in potassium phosphate buffer at pH 7.4 using amorphous material, and LogD was measured in a miniaturized shake-flask assay using DMSO stock solutions (Table 3). $pK_a$ was determined potentiometrically, or calculated using MoKA v.2.6.5 for atazanavir and ritonavir, for which low solubilities prevented experimental $pK_a$ determination. All compounds except for telaprevir contained acidic and/or basic functionalities; seven were predominantly charged at a physiological pH of 7.4 and four were neutral. Seven of the 11 compounds had high solubilities (>100 μM) and all of them were charged and had lipophilicities in the preferred drug-like range ($LogD_{7.4}$ = appr. 1–3). The remaining four compounds had lower solubilities (<25 μM), most likely as a result of their higher lipophilicities ($LogD_{7.4}$ >3.8).

**Table 3.** Experimentally determined physicochemical properties for the bRo5 drugs and clinical candidates included in the training set.

| Class | Name | Solubility (SEM)[a] (μM) | $LogD_{7.4}$ (SEM)[b] | $pK_a$[c] | Ionization state at pH 7.4 |
|---|---|---|---|---|---|
| Erythronolide | Azithromycin | 1920 (135) | 1.1 (0.1) | 8.69, 9.45 | 2+ |
| Erythronolide | Clarithromycin | 746 (47) | 1.6 (0.1) | 9.10 | 1+ |
| Erythronolide | Erythromycin | 1405 (81) | 0.9 (0.4) | 8.87 | 1+ |
| Erythronolide | Roxythromycin | 1510 (24) | 1.8 (0.1) | 9.13 | 1+ |
| Erythronolide | Telithromycin | 1960 (141) | 2.1 (0.1) | 4.91, 8.69 | 1+ |
| HIV-1 inh. | Atazanavir | 2.4 (0.5) | 4.2 (0) | 5.01* | neutral |
| HIV-1 inh. | Ritonavir | 0.3 (0.1) | 4.6 (0.1) | 2.43*, 3.45* | neutral |
| HIV-1 inh. | Saquinavir | 24.5 (3.5) | 4.7 (0.2) | 7.16 | neutral |
| NS3/4A inh. | Asunaprevir | 160 (25) | 3.1 (0.1) | 5.7 | 1− |
| NS3/4A inh. | Telaprevir | 4.3 (0.1) | 3.8 (0.2) | - | neutral |
| Rifamycin | Rifampicin | 183 (8.0) | 1.3 (0.1) | 2.97, 7.50 | zwitterion |

[a]Standard error based on four repeats. [b]Relative standard error based on 3–6 repeats. [c]Measured by potentiometry; based on three repeats with SEM<0.03. Values marked with an asterisk (*) were calculated with MoKa v.2.6.5. ADMET Predictor v7.2 provided similar values, i.e. 4.63 for atazanavir and 2.47 and 4.46 ritonavir.

**Models for Predicting Cell Permeability and Solubility**. We assessed if 2D or 3D PSA measures could explain the balance between cell permeability and solubility for the bRo5 drugs in the training set, and if quantitative models could be derived for these properties. When calculating 3D PSA, the impact of including polarity originating from partially charged atoms was compared to the traditional approach focused on oxygen, nitrogen and their attached hydrogen atoms.[39] Throughout this analysis TPSA was calculated as described by Ertl,[41] using the implementation in InstantJChem. Different programs (MOE, CDK, Spartan and PyMol) and different methods for atom type selection and partial charge calculation (using PM3, MMFF and

Gasteiger methods) were initially explored for 3D PSA calculation. PyMol was selected because of its open-source accessibility and flexibility in defining surface area and atoms for the PSA calculation (commands available in SI). Thus, in the presented analyses PSA was calculated by three distinct approaches: i) 2D TPSA, assigning as polar atoms any N, O, or their attached H, ii) molecular 3D PSA (i.e. the surface area accessible to a probe of zero radius) and iii) solvent-accessible 3D PSA (i.e. the surface area accessible to a water-sized probe of 1.4 Å radius). The two 3D PSA methods were applied to all conformations, i.e. to all crystal structures, of each compound. Definitions that expanded the list of included atoms with atoms having PM3 calculated partial charges above specified thresholds were also investigated for the two 3D approaches—thereby also incorporating moderately polar atoms in the PSA (cf. next section). All regressions discussed can be found in section 2 of the Supporting Information along with $r^2$, leave-one-out $q^2$, $p$-values and RMSD of training and test sets.

Efflux-inhibited (passive) permeability ($P_{app}$ AB +inh) showed only a modest correlation with TPSA in our dataset ($r^2$=0.36, $p$=0.068; Figure 3A) and a moderate correlation with $LogD_{7.4}$ ($r^2$ = 0.63, $p$ = 0.0061, Figure S2). Use of *molecular* 3D PSAs improved the correlation with $P_{app}$ AB +inh slightly as compared to TPSA, whereas *solvent-accessible* 3D PSAs gave the best correlations (Figure S3A). Notably, the correlation obtained using solvent-accessible 3D PSA with the traditional definition of polar atoms (i.e. N, O and attached H) ($r^2$=0.26, p=0.14) was significantly improved by including moderately polar atoms when calculating PSA. An optimal correlation was obtained when atoms with absolute partial charges >0.6 were included as polar, giving an $r^2$ of 0.90 ($p$=3.1×$10^{-5}$) for solvent-accessible 3D PSA (Figure 3B). The statistical validity of the correlation was demonstrated by a permutation procedure, in which the order of

the dataset was randomly shuffled 20,000 times (i.e., each permeability measurement was randomly associated with one PSA value from the dataset). In no case did the resulting permuted $r^2$ reach the observed value of 0.90; this corresponds to an empirical $p < 5\times10^{-5}$.

Importantly, the minimum solvent-accessible 3D PSA for each compound was consistently better correlated with efflux-inhibited permeability (log $P_{app}$ AB +inh) than were the corresponding maximum or average PSAs, regardless of the partial charge threshold (Figure S3A). This observation is consistent with conformation-dependent shielding of polar functional groups as the compounds pass through the lipophilic cell membrane interior. It also indicates that solvent-accessible 3D PSA and correct polar atom selection are important factors to achieve optimal predictive power for passive cell permeability in bRo5 space.

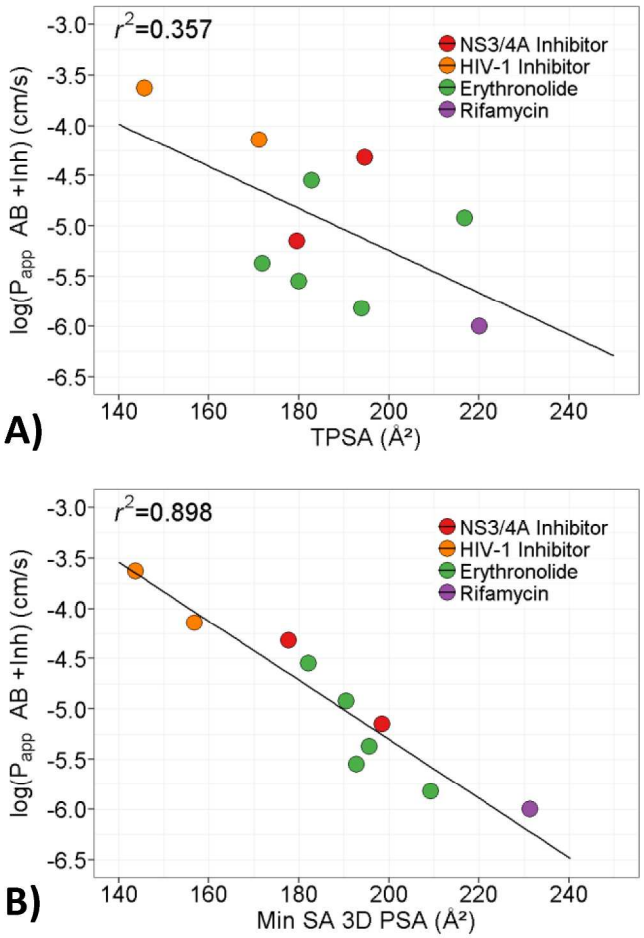



**Figure 3.** Permeability of the training set compounds [log ($P_{app}$ AB +inh) cm/s] and its correlation to A) topological PSA (TPSA, Å$^2$) and B) minimum solvent-accessible 3D PSA (Min SA 3D PSA, Å$^2$) calculated from N, O and attached H atoms with inclusion of atoms with absolute partial charges >0.6 (as calculated by the PM3 method) as polar. Models were derived from the minimum and maximum solvent-accessible 3D PSA across all conformations for the training set drugs. Refer to Figure S3 for models from alternative PSA definitions, including molecular and solvent accessible 3D PSA that includes atoms based on other partial charge thresholds. ***Correlations***: $\log(P_{app}\ AB\ +\ Inh) = -0.02098 \times TPSA - 1.0492$ , $r^2 = 0.36$, $p = 0.068$, LOO $q^2 = 0.071$ and $\log(P_{app}\ AB\ +\ Inh) = -0.02943 \times Min\ SA\ 3D\ PSA + 0.5825$ (atoms with absolute partial charges >0.6 were included as polar), $r^2 = 0.90$, $p = 3.1 \times 10^{-5}$, LOO $q^2 = 0.85$.

The use of minimally exposed polarity in explaining permeability bRo5 was further validated using the eight drugs and clinical candidates in the external test set (Table 1, Supporting Table S3). In line with the observations from the training compounds, the minimum solvent-accessible 3D PSA gave the best predictions for the efflux-inhibited cell permeabilities of the compounds in the test set (Figure 4, RMSE = 0.71, corresponding to 5.1-fold average error). The largest deviation was seen for paclitaxel, which is attributable to that its minimum solvent-accessible PSA was outside the range seen in the training data. Excluding paclitaxel decreased the average error to 0.59 (3.9-fold), comparable to the predictivity of more complex structure-permeability models in bRo5 space.[26] Notably, the two bRo5 peptides cyclosporine A and actinomycin D differed from the training compounds by being substantially larger (MW = 1203 and 1255 Da, in contrast to 680–840 Da in the training set) and containing more polar atoms (TPSA = 279 and 356 $Å^2$, compared to 146–220 $Å^2$). Still, permeability was well predicted for both cyclic peptides by their minimally exposed solvent-accessibly 3D PSA (1.2 and 1.6-fold deviations from the average efflux-inhibited permeability reported).

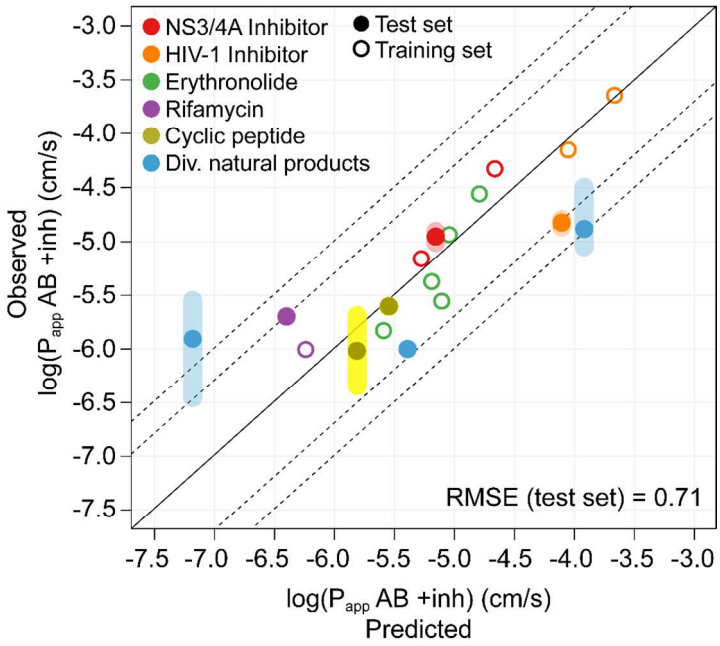



**Figure 4.** Permeability predictions for the external test set. Predictions are based on the correlation between efflux-inhibited Caco-2 cell permeability and solvent-accessible 3D PSA for the compounds in the training set (hollow circles, c.f. correlation in legend of Figure 3B). This correlation was used to predict the permeability for the external test set (solid circles). Coloured bars for the test set indicate the range of observed permeabilities for these compounds. The root mean squared error of prediction (RMSE) for the test set was 0.71 (5.1-fold). This was reduced to 0.59 (3.9-fold) if paclitaxel was excluded, the PSA of which was outside the range in the training data.

In comparison to cell permeability, aqueous solubility was somewhat better explained by TPSA ($r^2$ = 0.53, $p$ = 0.01; Figure 5A). Again, the correlation improved substantially when the three-dimensional structure was taken into account. As for permeability, the optimal PSA definition included moderately polar atoms (absolute partial charges >0.5, Figure 5B), further supporting a

role of these in interactions with the surrounding medium. For solubility, however, much smaller differences were obtained when the minimum or maximum 3D PSA for the compounds in the training set was used ($r^2$ = 0.83, $p < 1\times10^{-3}$, Figure S3B). Also, molecular surface areas gave better correlations than solvent-accessible areas. Since the former are less sensitive to conformational variation, the improved correlation suggests that the overall polarity in the molecule, rather than a single, specific conformation is the most predictive for solubility. These observations may therefore reflect a more extensive conformational sampling in aqueous media than in a lipid membrane environment. This hypothesis is tentatively supported by studies of two of the very few well characterized molecular chameleons, i.e. cyclosporin A[54] and a de novo designed macrocycle,[26] both of which have been found to display such environment dependent conformational flexibility. Similar to the permeability model above, the probability of obtaining a solubility model with $r^2 \geq 0.83$ using randomly permuted data was low ($p$ = 0.0003). A strong correlation was observed between aqueous solubility and experimentally determined $logD_{7.4}$ ($r^2$ = 0.82, $p$=1.2×10$^{-4}$; Figure 5C). Similarly, adding a calculated lipophilicity descriptor (cLogP) to the model based on the optimal calculated 3D PSA improved the correlation ($r^2$ = 0.90, $p$=8.4×10$^{-5}$, Figure S4B).

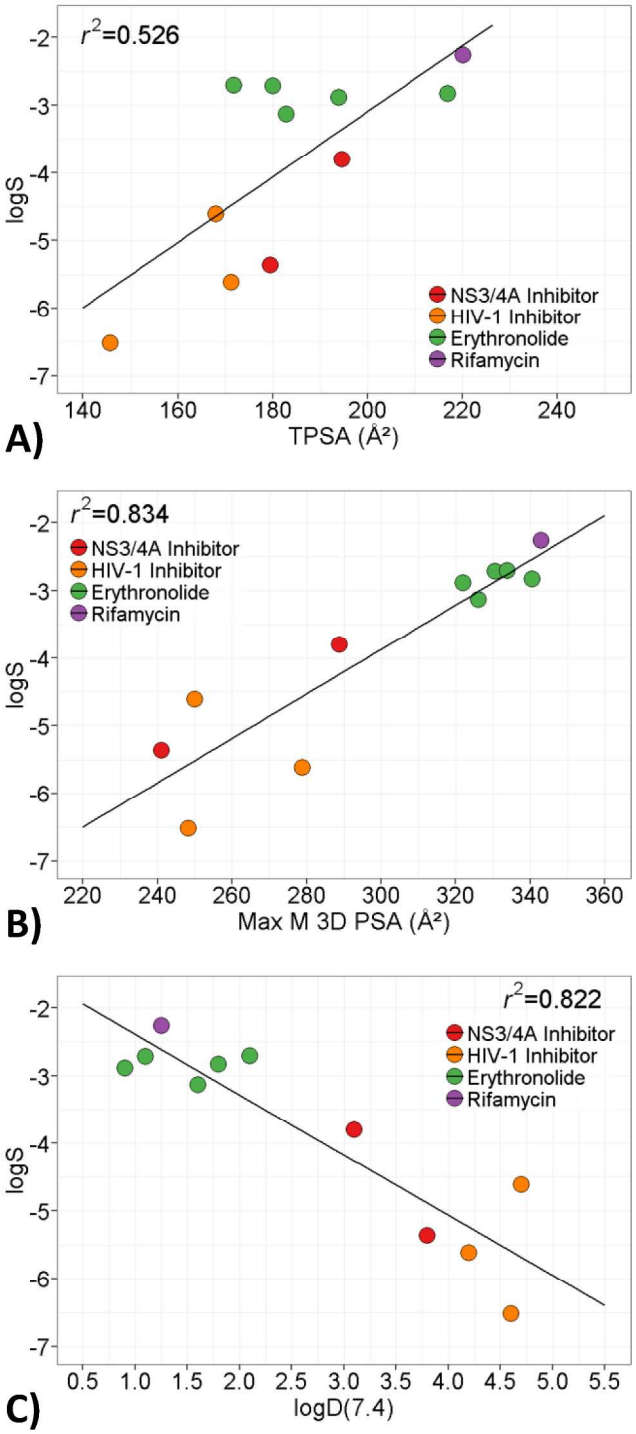



**Figure 5.** Solubility (logS) and its correlation to A) topological PSA (TPSA, Å$^2$) and B) maximum molecular 3D PSA (Max M 3D PSA, Å$^2$) calculated from N, O and attached H atoms with inclusion of atoms with absolute partial charges >0.5 (as calculated by the PM3 method) as

polar and C) experimental $LogD_{7.4}$. Models were derived from the minimum and maximum molecular 3D PSA across all conformations for the 11 selected drugs. Refer to Figure S3 for models from alternative PSA definitions, including molecular and solvent-accessible 3D PSA that includes atoms based on other partial charge thresholds ***Correlations:*** $logS = 0.04836 \times TPSA - 12.7623$, $r^2 = 0.53$, p-value = 0.012, LOO $q^2 = 0.35$ and $logS = 0.03290 \times Max\ M\ 3D\ PSA - NOH+0.5 - 13.738$, $r^2 = 0.83$, p-value = 8.66e-5, LOO $q^2 = 0.74$ and $logS = -0.8912 \times logD(7.4) - 1.4975$, $r^2 = 0.82$, p-value = 1.17e-4, LOO $q^2 = 0.71$.

**Compound- and Conformation-Dependent Variation in Molecular Properties.** Multiple sequential processes affect the permeability of drugs across cell membranes. Desolvation occurs as the drug leaves the extracellular aqueous environment and is followed by interactions with phospholipid head groups before it penetrates into the hydrophobic membrane interior. Then a similar but reversed sequence of events take place as the drug enters the cytosol. Each of these steps are likely differently affected by the drug's molecular properties. We were therefore intrigued by the fact that such a strong correlation to cell permeability was obtained using the minimally solvent-exposed 3D PSA as a single descriptor.

For example, the molecular radius of gyration and other descriptors reflecting the cross-sectional area of the permeant has been shown to be important factors in membrane permeability,[42, 55, 56] presumably reflecting the cost of forming cavities as the drug penetrates the phospholipid membrane. However, in our dataset, the radius of gyration was constrained to a relatively narrow interval (4.8–5.9 and 4.8–6.1 Å in the minimum-radius conformations of the training and test set compounds, respectively), suggesting a similar energetic cost for cavity formation (Figure S5A). Most compounds also displayed a relatively small variation between conformations, with the

maximum radii being, on average, 1.06 times larger than the minimum ones. The total solvent-accessible surface areas were also relatively similar between compounds (898–1024 and 937–1317 $Å^2$ in the minimum-surface area conformations for the training and test set, respectively) and between conformations of these (average fold-difference of 1.05 between maximum and minimum conformations, Figure S5B).

In contrast, the minimum solvent-accessible 3D PSA varied more between compounds (143–231 and 153–263 $Å^2$ for the training and test set conformations, respectively, Table S2 and Figure S5B). Also, PSA was more conformation-dependent, on average displaying 1.2-fold differences in exposed PSA. Differences in PSA ranged from quite small for roxithromycin and indinavir (ΔPSA: 7 and 13 $Å^2$, respectively) to very large for rifampicin, telithromycin, actinomycin D, faldaprevir and cyclosporin A (ΔPSA: 59, 60, 62, 72, and 79 $Å^2$). The average difference of 37 $Å^2$ in our dataset roughly corresponds to a shielding of 3–4 polar atoms in the minimum as compared to the maximum solvent-accessible conformation.

The impact of dynamically exposed 3D PSA that we observe in the present dataset potentially arises from the dual role of polar interactions in desolvation and resolvation and in the energy barrier for the drugs to cross the hydrophobic membrane interior. Compounds that have multiple energetically favorable conformations may thus attain lower desolvation costs and lower penetration barriers by adopting the most favorable (i.e. the least polar) conformation when permeating cell membranes. Although this cannot be decisively concluded, the fact that our dataset samples from 7 of 11 major structural classes in which orally bioavailable bRo5 drugs have been categorized suggests that such chameleonic behavior may be a general feature of drugs in the oral bRo5 subspace.

**Influence of Polar Atom Selection on PSA.** Interestingly, the optimal polar atom selection threshold for modeling permeability and solubility included atoms with absolute partial charges >0.6 and >0.5, respectively. This resulted in some atoms that would not normally be selected as polar being included in the polar surface area, for instance atoms adjacent to polar groups but also some terminal methyl carbon atoms (but not their attached hydrogens). A logical explanation for these findings is that at very low thresholds, polar atoms in addition to the typical N, O and attached H are taken into account, but also atoms which are not genuinely polar, thereby reducing the correlations. At high thresholds (absolute partial charge >0.8), some genuinely polar atoms are excluded, again decreasing the correlation to permeability and solubility. This also manifests as a loss of the conformation-dependence in the PSA-permeability relationship, i.e. minimum and maximum PSA show equally poor correlation with permeability (Figure S3). The partial charge threshold affected the PSA for most compounds in the dataset, with differences ranging from 15 to 73 $Å^2$ (mean ± SD: 46 ±20 $Å^2$) between the least inclusive PSA definition (N, O and attached H) and the optimal partial-charge threshold of >0.6 for minimum or maximum solvent-accessible 3D PSAs. Our results thus suggest that not only the conformation and orientation (exposure) of polar atoms, but also the different distribution of charge on atoms between conformations is important for accurate prediction of permeability and solubility.

**Structural Consequences of Conformational Flexibility.** The crystal structures presenting minimum and maximum solvent-accessible 3D PSA for each compound were systematically analyzed to determine what substructural features drive these conformational differences.

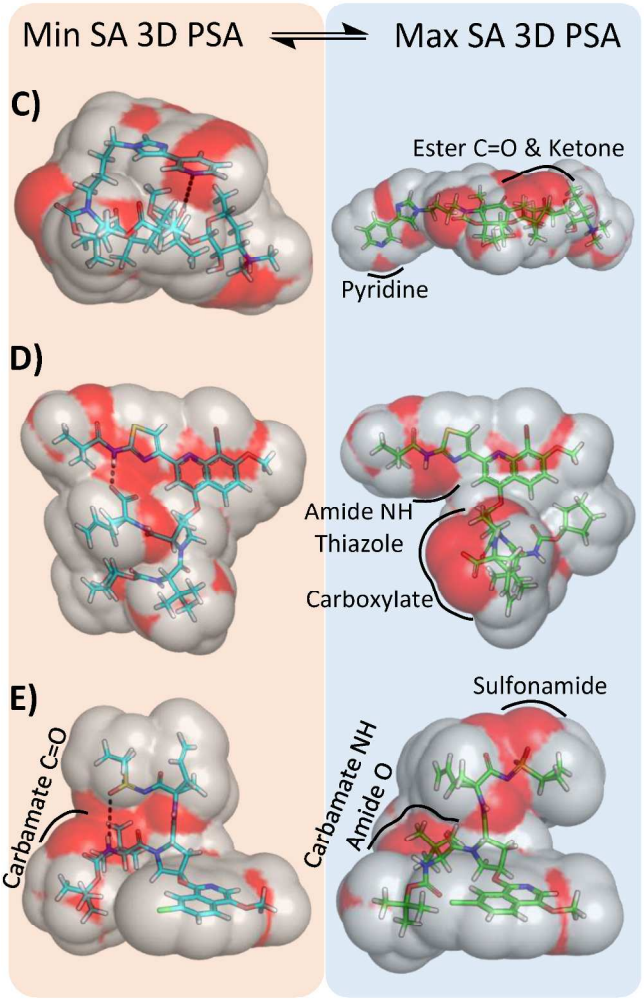
Conformational changes occurred in the core skeleton for half of the 24 compounds in our crystal structure dataset, and most compounds had changes in the orientation of their side chains (Figure 6A). Most conformational variation originated from freely rotatable bonds (bonds directly connected to an amide are not included), both for compounds with conformational changes in the core and for those with changes in the side chains. Additionally, eight compounds displayed flexibility involving bonds directly connected to an amide, while cycloalkane moieties changed conformation in three compounds. For example, telaprevir has an *N*-cyclopropyl amide that undergoes rotation of both its linked bonds, resulting in the amide oxygen's solvent-accessible 3D PSA as well as that of an adjacent carbonyl group becoming buried. Two of the three compounds that buried 3D PSA through conformational changes in cycloalkane moieties were from the rifamycin class: rifampicin forms an IMHB when its piperazine group adopts a boat-like conformation, and in rifabutin the piperidine NH is occluded by a neighboring bulky lipophilic side chain when the piperidine adopts an alternate chair-like conformation. Thus, even though cycloalkyl groups and amides are semi-rigid functional groups, they can still adopt multiple stable, low-energy conformations with different solvent-accessible 3D PSA. It may be speculated that this semi-rigidity might make these and other semi rigid groups particularly useful in design of multi-conformer "chameleonic" compounds.

**A)**

| Region of change | Occurrence |
|---|---|
| Core | 12/24, 50 % |
| Side chain | 19/24, 79 % |
| Freely rotatable bond (core) | 9/24, 38 % |
| Freely rotatable bond (side chain) | 13/24, 54 % |
| Amide and adjacent bonds | 8/24, 33 % |
| Cycloalkane | 3/24, 13 % |

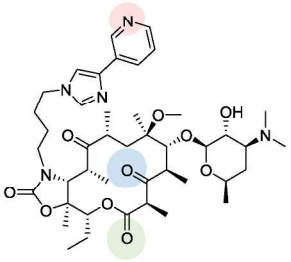
**B)**

| Interaction | Occurrence | Average and range of SA 3D PSA change |
|---|---|---|
| Aryl shielding | 7/24, 29 % | 28 Å$^2$ (13-55 Å$^2$) |
| Alkyl sheilding | 6/24, 25 % | 14 Å$^2$ (7-20 Å$^2$) |
| IMHB | 8/24, 33 % | 27 Å$^2$ (9-77 Å$^2$) |



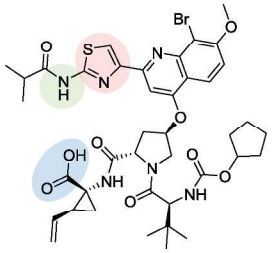

**Telithromycin**

| | min | max |
|---|---|---|
| Pyridine N | 16 | 25 |
| Ester C=O | 0 | 25 |
| Ketone | 2 | 23 |
| Total 3D PSA | 196 | 256 |



**Faldaprevir**

| | min | max |
|---|---|---|
| Carboxylate | 16 | 77 |
| Amide NH | 2 | 16 |
| Thiazole | 14 | 19 |
| Total 3D PSA | 194 | 267 |



**Asunaprevir**

| | min | max |
|---|---|---|
| Sulfonamide | 41 | 63 |
| Amide O | 8 | 11 |
| Carbamate NH | 2 | 8 |
| Carbamate C=O | 28 | 10 |
| Total 3D PSA | 178 | 196 |

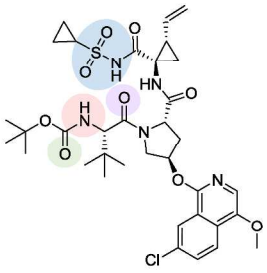


**Figure 6.** Conformational changes in crystal structures that display minimum and maximum solvent-accessible 3D polar surface area (Min and Max SA 3D PSA). A) Summary of substructural regions from which flexibility originates in the 24 investigated drugs and clinical candidates. B) Interactions formed in the minimum PSA structures as compared to those displayed in the maximum PSA structure. Crystal structures of C) telithromycin (1YIJ and

1P9X), D) faldaprevir (MEBYEZ and 3P8N) and E) asunaprevir (MIYWOI and 4WH6) with conformations shown that display minimum (orange box) and maximum (blue box) solvent-accessible 3D PSA. The structures are shown as sticks (carbons blue/green, nitrogens dark blue, oxygens red, sulfurs yellow and hydrogens white) with solvent-accessible surface area shown as red (polar) and white (non-polar). Solvent-accessible 3D PSA (SA 3D PSA, $Å^2$) was calculated from N, O and attached H atoms with inclusion of atoms with absolute partial charges >0.6 (as calculated by the PM3 method) as polar. Polar and IMHB interactions are shown as dotted black lines. Polar functional groups that are differently exposed depending on conformation (leading to differences in 3D PSA) are outlined and labeled in the 3D molecular structure, detailed in the accompanying tables and shaded in the 2D structures.

Bulky side chains can be used to hide adjacent PSA in a non-polar environment, but this may lead to reduced solubility,[20, 23, 28] unless conformational flexibility allows re-exposure of the hidden PSA in aqueous environments.[22, 26] In half of the compounds in our crystal structure dataset, one or more bulky side chains occluded polar functionalities in the minimum 3D PSA conformation, but not in the corresponding maximum PSA conformation (Figure 6B). Six of the affected compounds were from the de novo designed HIV and HCV protease inhibitor series, while the remaining six were in the natural product derived and cyclic peptide classes. The largest change in PSA was observed in telithromycin, in which the aromatic side chain folds over the polar macrocyclic core and shields a significant amount of solvent-accessible 3D PSA (55 $Å^2$, Figure 6C).

Flexibly attached aromatic rings such as the one in telithromycin accounted for half of all such events of dynamic shielding of PSA, and lead to occlusion of an average of 28 $Å^2$ of solvent-

accessible 3D PSA. Alkyl shielding was equally common but shielded less PSA (average of 14 Å$^2$). Alkyl shielding is best exemplified by faldaprevir's isopropyl and vinylcyclopropyl side chains (Figure 6D). Though relatively rigid, upon formation of a charge-reinforced IMHB, these alkyl groups contribute to further reductions in 3D PSA. Interestingly, most instances of aryl shielding (5/7) came from the *de novo* designed HIV-1 and HCV NS3/4A protease inhibitors, and the remaining two instances came from the designed aryl side chains of telithromycin and paclitaxel. Alkyl shielding came predominantly from natural product and cyclic peptide classes. We hypothesize that the higher shielding of PSA by aromatic groups is partly due to their size and rigidity, but also to the ability of aromatic groups to participate in weak dipole or weak IMHB interactions with polar groups that can stabilize their conformational shielding.[57] Preliminary support for this can be seen in telithromycin, where the flexible aromatic side chain folds directly over the ketone and a weak dipole–π-interaction can be formed (Figure 6C). Flexibly linked aromatic side chains may therefore be particularly useful in design of "chameleonic" compounds that display a dynamic solvent-accessible 3D PSA when adapting to different environments.

Formation of IMHBs is another theme that has been highlighted for reduction of PSA and a subsequent increase of cell permeability.[29-36] Motifs for formation of five to eight-membered IMHB pseudoring systems have been derived by exhaustive analysis of crystal structure databases and may serve as inspiration in drug design.[34] However, as noted, IMHBs that are stable in both polar and non-polar environments (static IMHBs) can reduce solubility in aqueous environments.[20, 27, 28, 34] Indeed, just under half (10/24) of the drugs and clinical candidates investigated herein made 1-4 IMHBs that were unchanged between the conformations exposing the minimum and maximum solvent-accessible 3D PSA. All compounds with such stable

IMHBs were from the natural product and cyclic peptide classes. For drug design purposes, dynamic—as opposed to static—IMHBs are of greater interest as they potentially allow for both high solubility and permeability.[28, 30, 34, 58] Eight of the 24 analyzed compounds formed additional IMHBs in their minimum solvent-accessible 3D PSA conformation that were not formed in the maximum PSA conformation. Six of these compounds were from the natural product classes, while two were de novo designed drugs. This indicates that natural products may have evolved to take advantage of both static and dynamic IMHBs for reduction of compound polarity, but also that medicinal chemists have not yet capitalized on dynamic IMHBs to their full potential when designing drug candidates in bRo5 space.

The average solvent-accessible 3D PSA hidden by dynamic IMHBs was 27 Å$^2$, but this number varied considerably from 9 to 77 Å$^2$ between different compounds (Figure 6B). The surprisingly large PSA reduction of 77 Å$^2$ originates from faldaprevir, which makes a charge-reinforced IMHB between its carboxylate and isobutyramide NH (Figure 6D). Faldaprevir displays an interesting synergy between the two most prevalent methods for burying solvent-accessible 3D PSA, as the two polar groups are highly exposed in the maximum 3D PSA conformation but shielded by the bulky side chains in the hydrogen bonded minimum PSA conformation.

When examining the substructural motifs present in these dynamic shielding events, a diverse range of functional groups as well as sizes and distances between the groups was seen. Dynamic IMHBs formed pseudorings in eight compounds, where all but three of the eleven instances had 10-17 atoms in the pseudoring. This indicates opportunities to extend the use of dynamic IMHBs in design beyond motifs based on five to eight-membered pseudorings,[34] many of which are stable in different environments. Amide NH (5/11 instances) and amide CO (4/11 instances) were the most common donors and acceptors. All but one of the dynamically formed IMHBs

contained two or more rotatable bonds, predominantly centered around $sp^3$-hybridized carbon atoms within the IMHB ring. In addition, all dynamically formed IMHB contained rigidifying functional groups such as amides, alkyl rings, aryl rings, esters and carbon-carbon double bonds. There is no doubt that a subtle play between rigid and flexible elements is required to obtain dynamic, environment-dependent IMHBs. The most common motif in aryl shielding events involved 6-atom linkers between the polar group and the aryl ring (3/7 instances), with the remainder of the events ranging between 4 and 15 atoms in the linker. Phenyl groups accounted for the majority of aryl groups (6/7 aryl groups), probably due to their prevalence in the dataset, while polar groups were more diverse and included carbamates, hydroxyl groups, amides, sulfonamides and ketones. Finally, dynamic IMHBs were equally frequent between side chains as within cores (each 5/11 instances), while aryl shielding of polar groups were more frequent between side chains than side chain to core.

**Predicting the Effect of Dynamic Polarity Exposure on Permeability.** The average effect of the different intramolecular interactions on polar surface area—when combined with our model correlating passive cell permeability to solvent accessible 3D PSA—provide an opportunity to predict the average impact from each intramolecular interaction on passive cell permeability. Accordingly, aryl shielding and IMHB formation led to an average of 28 and 27 $Å^2$ reduction in solvent accessible 3D PSA, respectively, which is predicted to yield ~6-7 fold improvement in permeability. Alkyl shielding, on the other hand, had a much lower average effect on 3D PSA of 14 $Å^2$, and is expected to increase passive permeability ~2-3 fold.

The largest variations in solvent accessible 3D PSA were observed for cyclosporin A (79 $Å^2$) and faldaprevir (77 $Å^2$), both of which were included in the external test set. Notably, their minimum

solvent accessible 3D PSA (208 and 194 Å$^2$, respectively) corresponds to predicted Caco-2 cell permeabilities of $2.9 \times 10^{-6}$ and $7.5 \times 10^{-6}$ cm/s, which both compare favorably to experimental Caco-2 data ($2.5 \times 10^{-6}$ cm/s for cyclosporin A; $8–11 \times 10^{-6}$ cm/s in two independent studies for faldaprevir[40, 59]). The >75 Å$^2$ lower solvent accessible 3D PSAs of these compounds' least polar conformations, as compared to their most polar ones, corresponds to a predicted >180-fold improvement in permeability. This may well explain why the observed permeabilities are much greater than would be expected from the 2D TPSA for these structurally very different compounds, and suggests a considerable potential of incorporating dynamic IMHBs and aryl/alkyl shielding in compound design. In addition, the correlation between permeability and solvent-accessible 3D PSA allows calculation of upper limits of 3D SA PSA that are compatible with satisfactory cell permeabilities. Thus, moderate-to-high cell permeabilities ($P_{app}$ AB +Inh ranging from 1 to $10 \times 10^{-6}$ cm/s) that are predictive of medium-to-high absorption from the intestine,[60] translate into solvent accessible PSAs ranging from ≤220 to ≤190 Å$^2$ when atoms with partial charges >0.6 are included in the PSA definition.

## CONCLUSIONS

Our analysis of drugs in the DrugBank database revealed a trend that PSA is buried in minimum energy conformations when MWs exceed 500-700 Da. This indicates that conformation dependent PSA is a common feature of larger drugs, which may allow them to behave as molecular chameleons that adapt their properties to their environment.[28, 30, 32, 33] The potential for conformational flexibility to provide chameleonic properties was further investigated by analysis of 24 orally administered drugs and clinical candidates, 19 of which had properties residing in bRo5 space. Examination of multiple crystal structures for each compound revealed that polar

functionality was buried to various extents and that solvent accessible 3D PSA varied by up to 79 $Å^2$ between conformations.

The minimum solvent-accessible 3D PSA calculated from the crystal structures of the drugs and clinical candidates in our training set displayed an excellent correlation with passive cell permeability. This correlation also applied to an external test set of additional drugs and clinical candidates in bRo5 space. Altogether, these findings suggest that the minimum solvent-accessible 3D PSA for relevant conformations can be used to predict cell permeability in bRo5 space, *provided that* conformations in a lipid-like environment can be calculated with reasonable accuracy. Accurate procedures for conformational sampling of macrocyclic peptides were reported most recently,[61] which provides hope that breakthroughs may also be made for conformational sampling of druglike compounds in bRo5 space in the near future. Solubility, which is also crucial for oral administration, was found to be less conformation-dependent, with similar correlations for minimum, average and maximum exposed PSA. This may reflect a greater conformational flexibility in aqueous than in lipid membrane environments, as has been suggested for cyclosporin A.[54] Importantly, correlations for both permeability and solubility were significantly improved when partially charged atoms were included in the calculations of PSA, as compared to predictions based solely on nitrogen, oxygen and attached hydrogen atoms.

Current knowledge of how to incorporate conformational flexibility in design of chameleonic drugs is limited. Our inspection of crystal structures suggests that approaches based on incorporation of flexibility in attached side chains of macrocycles as well as non-macrocycles are more likely to be successful in providing chameleonic properties than attempts to adjust flexibility in the backbone. In particular, flexibly linked aromatic side chains have been employed successfully in de novo designed drugs, but also in modification of natural products.

Environment-dependent, dynamic IMHBs were more frequent in the natural products studied herein than in the de novo designed drugs. The finding that most dynamic IMHBs observed in our dataset involved pseudorings with more than ten members indicates a scope for utilization in compound design beyond the five to eight-membered pseudoring motifs already highlighted by others.[34] Incorporation of dynamic IMHBs and flexibly linked aromatic side chains was predicted to improve permeability by ~6-7 fold per effective interaction, whereas aliphatic side chains only provide a ~2-3 fold improvement. Significantly greater improvements may be obtained, as illustrated by the charge-reinforced IMHB of faldaprevir. However, if intramolecular interactions are not flexibly formed, any improvement in permeability is likely to be accompanied by reduced solubility.[20, 27, 28]

In conclusion, it is becoming increasingly clear that traditional 2D descriptors such as TPSA, HBA and HBD fail to effectively capture the properties that impart cell permeability and oral absorption to chameleonic compounds in bRo5 space. Instead, it appears that conformational preferences and flexibility must be considered, and that use of 3D descriptors provides significant advantages.[32, 33] 3D descriptors such as the radius of gyration ($R_{gyr}$),[42] the degree of IMHB formation,[62] and free energies of desolvation[63] have showed promise, and a procedure relying on a combination of QSPR modeling and assessment of the polarity of conformational ensembles was recently presented.[26] As described herein, the minimum solvent-accessible 3D PSA gave the best correlation to cell permeability when moderately polar atoms were included in the PSA calculation. Compounds that adopt one or several conformations that expose less than 190-220 $Å^2$ solvent-accessible 3D PSA are likely to display medium-to-high oral absorption, provided that the compounds are not associated with major transporter-mediated efflux. We are therefore optimistic that progress in conformational sampling of compounds in bRo5 space, e.g.

as recently demonstrated for macrocyclic peptides,[61] in combination with introduction of more informative 3D descriptors will continue to improve our insight into what determines drug-like properties in this space. Rational design also of non-peptidic drugs in bRo5 space that can be administered orally may thus become a reality in a not too distant future.

## EXPERIMENTAL SECTION

**Source of Drugs and Clinical Candidates.** Atazanavir sulfate, azithromycin, clarithromycin, erythromycin, roxithromycin, ritonavir and telaprevir were purchased from Selleckchem. Asunaprevir was from Medchemtronica, rifampicin from Sigma, saquinavir from MedChem Express, and telithromycin from TOKU-E, respectively. All compounds had a purity >95.8%.

**Calculating PSA for Drugs in DrugBank.** The DrugBank dataset[40] was downloaded on 22/01/2017 and filtered to contain drugs with MW 100 Da – 2000 Da. TPSA and molecular 3D PSA were calculated in InstantJChem v6.2.0.953 and Pymol 1.7.0.1, respectively, while low energy 3D conformers were calculated in Corina v3.2.

**3D Conformer Preparation from Experimental Data.** The IUPAC names, SMILES and common synonyms were obtained for the 24 compounds from the ChemSpider database (www.chemspider.com). All instances of crystal structure data for the 24 compounds were extracted from the PDB (www.rcsb.org/pdb) and CSD (www.ccdc.cam.ac.uk), using searches by common name, synonyms, and chemical structure. From the PDB, only crystal structures with a resolution < 3.5 Å were considered. They were inspected visually to ensure that all conformers that were included had well defined electron densities for the drug or clinical candidate.

Conformers were also check to ensure that there were no steric clashes and suitable geometry, occupancies, and temperature factors relative to the bound protein structure to ensure that modelled conformations were reliable. The structures were initially analyzed in MOE v 2014.10 (Chemical Computing Group, www.chemcomp.com) and hydrogen atoms were added according to the measured or predicted major ionization state at pH 7.4. The structures were relaxed in the MMFF94x force field with a Born implicit electrostatic model ($\varepsilon = 80$), with maximum deviation from the original structures set to RMSD $\leq$0.5 Å. Partial charges were then calculated in Spartan 14 v1.1.4 using the PM3 semi-empirical method. The structures were aligned by heavy atoms using MOE and pairwise heavy atom RMSD was calculated using the mol_rmsd svl plugin. Conformers were clustered using a heavy atom RMSD threshold of 0.75 Å.

**Molecular Descriptor Calculations** Topological polar surface areas (TPSA) were calculated from SMILES structural representation using the method of Ertl et al.[41], as implemented in InstantJChem v6.2. cLogP was also calculated in InstantJChem v6.2. Three-dimensional PSA was calculated in PyMol v1.7.0.1, using molecular surface areas calculated from atomic van der Waal's radii, or solvent-accessible surface areas calculated using a solvent molecule radius of 1.4 Å. Atoms were assigned as 'polar' either based solely on atom type (O, N, and attached H), or by also including atoms with absolute partial charges from the PM3 semi-empirical method above a defined threshold. Three dimensional PSA was calculated for all crystal structures of each of the 24 compounds in the crystal structure dataset. Further details, settings and commands are available in the supporting information.

**Caco-2 Cell Permeability Assay.** Caco-2 cells (originally obtained from the American Type Culture Collection; ATCC), passage 95-105, were maintained in Dulbecco's modified Eagle's medium (DMEM), containing 10 % fetal calf serum and 1 % non-essential amino acids. Cells were cultured on Transwell polycarbonate filters (diameter 12 mm, pore size 0.4 μm) as previously described.[64] Briefly, cells were seeded at a density of $0.5 \times 10^6$ cells/filter and maintained in DMEM with 10 % fetal calf serum, 1 % non-essential amino acids, 100 units/ml penicillin and 100 μg/ml streptomycin, for 21 days prior to permeability experiments.

Permeability measurements were performed as previously described.[64] Briefly, culture medium was replaced with preheated Hank's Balanced Salt Solution (HBSS) buffered with HEPES to pH 7.4, 30 minutes prior to the start of the experiment. For measurements of efflux-inhibited permeability, a cocktail of 50 μM quinidine, 20 μM sulfasalazine and 30 μM benzbromarone was included in the preincubation cocktail, and also in the donor and receiver solutions during the permeability experiment.

Apparent permeability was measured, at a compound concentration of 1-2 μM, in both apical-to-basolateral (A-B) and basolateral-to-apical (B-A) direction at pH 7.4 using a shaking speed of 500 rpm. Immediately after the start of the experiment a sample was removed from the donor compartment ($C_0$), and subsequent samples were taken from the receiver compartment at 15, 30 and 60 min. At the end of the experiment (t=60 min), a sample was removed from the donor chamber ($C_f$) for mass balance calculation. Compound concentrations were analyzed by LC-MS/MS (see Analytical procedure below) using eight-point standard curves. Values reported in Table 2 are means of three repeats, except for saquinavir and ritonavir (n=6, because of excessive mass balance in the uninhibited experiment).

The apparent permeability coefficient ($P_{app}$) of the compounds was calculated as:

$$P_{app} = \frac{{}^{dQ}/_{dt}}{C_0 \times A}$$

where dQ/dt is the linear rate of appearance of drug in the receiver compartment, A is the surface area of the filter, and $C_0$ is the initial drug concentration in the donor compartment. Efflux ratios (ERs) were calculated from the permeabilities in the basolateral-to-apical (BA) and apical-to-basolateral (AB) directions as:

$$ER = \frac{P_{app}\ BA}{P_{app}\ AB}$$

To confirm the integrity of the monolayer, the permeability of [$^{14}$C]mannitol was determined in parallel to the test compound, and the trans-epithelial electrical resistance (TEER) was monitored before and after each permeability experiment.

**Solubility Measurements.** Solubility was determined in 100 mM potassium phosphate buffer, pH 7.4. Solid material was weighed into glass HPLC vials and 400-500 µl potassium phosphate buffer was added, resulting in saturated solutions. Samples were equilibrated for at least 20 h at 37 °C in a heater-shaker at 900 rpm. After equilibration, samples were centrifuged at 10 000×g for 20 min to remove remaining particulate matter. The resulting supernatants were diluted 20- to 5000-fold in potassium phosphate buffer, and analyzed by LC-MS/MS using eight-point standard curves (see Analytical procedure below). Values reported in Table 3 are means of four determinations.

**Octanol-buffer Partitioning Measurements.** Octanol-buffer partition coefficients (logD) were determined using a miniaturized shake-flask procedure based on the traditional shake flask

technique.[65] 1-Octanol (HPLC grade ≥99 %, Sigma-Aldrich) and potassium phosphate buffer (KP) 10 mM, pH 7.4 (p.a. grade, Merck), were pre-saturated with each other by stirring overnight. Subsequently, the two layers were separated and stored for use in the assay. Two phase ratios (1:3 and 1:6 [KP:octanol]) were used, with total assay volume of 1.2 ml. The assay was started by pipetting 1 µl of each test compound into a 2 ml HPLC glass vial, to which octanol and KP buffer was added. The vials were sealed and shaken at 900 rpm in a V 2000 heater-shaker at ambient temperature (ca 23 °C) overnight. The vials were centrifuged at 3500 rpm for 20 min and then equilibrated in dark for an additional 24 h. Subsequently, the octanol phase was carefully separated from the aqueous phase using a pipette. Aliquots of the samples were diluted (125× and 2.5× for the octanol phase and aqueous phase, respectively). Compound concentrations in each phase were analyzed by LC-MS/MS using eight-point standard curves (see Analytical procedure below). Values reported in Table 3 are means of three to six repeats.

**Analytical Procedure.** Analysis was conducted on an Acquity UPLC coupled to a XEVO TQ triple-quadrupole mass spectrometer (both from Waters Corp., Milford, MA) with positive or negative electrospray ionization. For chromatographic separation, either a C18 BEH 1.7 µm or a HSS T3 1.8 µm column (both from Waters Corp.) was used, with a general gradient of 1 % to 90 % of mobile phase B over a total running time of 2 min. Mobile phase A consisted of 5 % acetonitrile and 0.1 % formic acid in purified water, and mobile phase B of 0.1 % formic acid in 100 % acetonitrile. The flow rate was set to 0.5 ml/min and 5 µL of the sample was injected. Warfarin was used as internal standard throughout the analysis. Mass spectrometric settings were optimized for each compound and are listed in Supplementary Table S4.

**Determination of pKa.** Acid dissociation constants were measured potentiometrically using a SiriusT3 instrument (Sirius Analytical Instruments) equipped with a Ag/AgCl double junction reference pH electrode and a turbidity sensing device. Titration experiments were conducted in 0.15 M KCl solution under nitrogen atmosphere at a temperature of 25 ± 1 °C. All tests were performed using standardized 0.5 M KOH and 0.5 M HCl as titration reagents. For low-solubility compounds, $psK_a$ values were determined in water-methanol mixtures (30-70 % methanol v/v) and extrapolated to pure aqueous conditions using the Yasuda-Shedlovsky method.[66] For compounds where experimental determination was precluded because of low aqueous solubility (atazanavir and ritonavir), or no sample (such as with the test set) dissociation constants were predicted with MoKA v.2.6.5. (www.moldiscovery.com). Reported experimental values are means of three determinations

## ASSOCIATED CONTENT

### Supporting information

The Supporting Information is available free of charge on the ACS Publications website.

Methods for calculation of 3D PSA. Models for permeability and solubility. Additional figures and tables referred to in the article.

## AUTHOR INFORMATION

### Corresponding Authors

*For P.M.: phone, +46 (0)18 4714630; e-mail, par.matsson@farmaci.uu.se.

*For J.K.: phone, +46 (0)18 4713801; e-mail, jan.kihlberg@kemi.uu.se.

### Author Contributions

M.R.S. and B.D. made equal contributions to the manuscript. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

**Notes**

The authors declare no competing financial interest.

## ACKNOWLEDGEMENTS

Support from the Magnus Bergvall and Åke Wiberg foundations, and the Swedish Fund for Research without Animal Experiments is gratefully acknowledged. The authors are also grateful to Ms. Elin Khan for assistance with permeability measurements and to ChemAxon for providing an academic license for Instant JChem.

## ABBREVIATIONS USED

AB, apical-to-basolateral; BA, basolateral-to-apical; BCRP, breast cancer resistance protein; bRo5, beyond rule of 5; CSD, Cambridge Structural Database; ER, efflux ratio; IMHB, intramolecular hydrogen bond; MRP2, multidrug-resistance associated protein 2; M 3D PSA, molecular 3D PSA; $P_{app}$, permability across Caco-2 cell monolayers; $R_{gyr}$, radius of gyration; SEM, standard error of the mean; TPSA, topological polar surface area.

## REFERENCES AND NOTES

1. Bunnage, M. E. Getting pharmaceutical R&D back on target. *Nat. Chem. Biol.* **2011,** *7*, 335-339.
2. Surade, S.; Blundell, T. L. Structural biology and drug discovery of difficult targets: The limits of ligandability. *Chem. Biol.* **2012,** *19*, 42-50.

3. Perola, E.; Herman, L.; Weiss, J. Development of a rule-based method for the assessment of protein druggability. *J. Chem. Inf. Model.* **2012,** *52*, 1027-38.
4. Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Disc.* **2002,** *1*, 727-730.
5. Rask-Andersen, M.; Masuram, S.; Schioth, H. B. The druggable genome: Evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. *Annu. Rev. Pharmacol. Toxicol.* **2014,** *54*, 9-26.
6. Giordanetto, F.; Kihlberg, J. Macrocyclic drugs and clinical candidates: What can medicinal chemists learn from their properties. *J. Med. Chem.* **2014,** *57*, 278−295.
7. Villar, E. A.; Beglov, D.; Chennamadhavuni, S.; Porco, J. A.; Kozakov, D.; Vajda, S.; Whitty, A. How proteins bind macrocycles. *Nat. Chem. Biol.* **2014,** *10*, 723-732.
8. Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. F. The exploration of macrocycles for drug discovery – an underexploited structural class. *Nat. Rev. Drug Discovery* **2008,** *7*, 608-624.
9. Marsault, E.; Peterson, M. L. Macrocycles are great cycles: Applications, opportunities, and challenges of synthetic macocycles in drug discovery *J. Med. Chem.* **2011,** *54*, 1961-2004.
10. Mallinson, J.; Collins, I. Macrocycles in new drug discovery. *Future Med. Chem.* **2012,** *4*, 1409-1438.
11. Doak, B. C.; Over, B.; Giordanetto, F.; Kihlberg, J. Oral druggable space beyond the rule of 5: Insights from drugs and clinical candidates. *Chem. Biol.* **2014,** *21*, 1115-42.
12. Doak, B. C.; Zheng, J.; Dobritzsch, D.; Kihlberg, J. How beyond rule of 5 drugs and clinical candidates bind to their targets. *J. Med. Chem.* **2016,** *59*, 2312–2327.
13. In reference 11 bRo5 space was defined as MW >500 Da, and at least one of MW 700-3000 Da, cLogP <0 or >7.5, HBD >5, HBA >10, PSA >200 Å2, or NRotB >20.
14. Doak, B. C.; Over, B.; Giordanetto, F.; Kihlberg, J. Oral druggable space beyond the rule of 5: Insights from drugs and clinical candidates. *Chem. Biol.* **2014,** *21*, 1115-1142.
15. Pye, C. R.; Hewitt, W. M.; Schwochert, J.; Haddad, T. D.; Townsend, C. E.; Etienne, L.; Lao, Y.; Limberakis, C.; Furukawa, A.; Mathiowetz, A. M.; Price, D. A.; Liras, S.; Lokey, R. S. Non--classical size dependence of permeation defines bounds for passive absorption of large drug molecules. *J. Med. Chem.* **2017,** *60*, 1665−1672.

16. DeGoey, D. A.; Chen, H.-J.; Cox, P. B.; Wendt, M. D. Beyond the rule of 5: Lessons learned from AbbVie's drugs and compound collection. *J. Med. Chem.* **2017,** *DOI: 10.1021/acs.jmedchem.7b00717*.

17. Rezai, T.; Yu, B.; Millhauser, G. L.; Jacobson, M. P.; Lokey, R. S. Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. *J. Am. Chem. Soc.* **2006,** *128*, 2510-2511.

18. Nielsen, D. S.; Hoang, H. N.; Lohman, R.-J.; Hill, T. A.; Lucke, A. J.; Craik, D. J.; Edmonds, D. J.; Griffith, D. A.; Rotter, C. J.; Ruggeri, R. B.; Price, D. A.; Liras, S.; Fairlie, D. P. Improving on nature: Making a cyclic heptapeptide orally bioavailable. *Angew. Chem. Int. Ed.* **2014,** *53*, 12059 –12063.

19. Wang, C. K.; Northfield, S. E.; Colless, B.; Chaousis, S.; Hamernig, I.; Lohman, R.-J.; Nielsen, D. S.; Schroeder, C. I.; Liras, S.; Price, D. A.; Fairlie, D. P.; Craik, D. J. Rational design and synthesis of an orally bioavailable peptide guided by NMR amide temperature coefficients. *Proc. Natl. Acad. Sci. USA* **2014,** *111*, 17504-17509.

20. Wang, C. K.; Northfield, S. E.; Swedberg, J. E.; Colless, B.; Chaousis, S.; Price, D. A.; Liras, S.; J., C. D. Exploring experimental and computational markers of cyclic peptides: Charting islands of permeability. *Eur. J. Med. Chem.* **2015,** *97*, 202-213.

21. Bockus, A. T.; Lexa, K. W.; Pye, C. R.; Kalgutkar, A. S.; Gardner, J. W.; Hund, K. C. R.; Hewitt, W. M.; Schwochert, J. A.; Glassey, E.; Price, D. A.; Mathiowetz, A. M.; Liras, S.; Jacobson, M. P.; Lokey, R. S. Probing the physicochemical boundaries of cell permeability and oral bioavailability in lipophilic macrocycles inspired by natural products. *J. Med. Chem.* **2015,** *58*, 4581−4589.

22. Bockus, A. T.; Schwochert, J. A.; Pye, C. R.; Townsend, C. E.; Sok, V.; Bednarek, M. A.; Lokey, R. S. Going out on a limb: Delineating the effects of β□branching, N□methylation, and side chain size on the passive permeability, solubility, and flexibility of sanguinamide A analogues. *J. Med. Chem.* **2015,** *58*, 7409−7418.

23. Thansandote, P.; Harris, R. M.; Dexter, H. L.; Simpson, G. L.; Pal, S.; Upton, R. J.; Valko, K. Improving the passive permeability of macrocyclic peptides: Balancing permeability with other physicochemical properties. *Bioorg. Med. Chem.* **2015,** *23*, 322-327.

24. Beck, J. G.; Chatterjee, J.; Laufer, B.; Kiran, M. U.; Frank, A. O.; Neubauer, S.; Ovadia, O.; Greenberg, S.; Gilon, C.; Hoffman, A.; Kessler, H. Intestinal permeability of cyclic

peptides: Common key backbone motifs identified. *J. Am. Chem. Soc.* **2012,** *134*, 12125-12133.

25. Ahlbach, C. L.; Lexa, K. W.; Bockus, A. T.; Chen, V.; Crews, P.; Jacobson, M. P.; Lokey, R. S. Beyond cyclosporine A: conformation-dependent passive membrane permeabilities of cyclic peptide natural products. *Future Med. Chem.* **2015,** *7*, 2121–2130.
26. Over, B.; Matsson, P.; Tyrchan, C.; Artursson, P.; Doak, B. C.; Foley, M. A.; Hilgendorf, C.; Johnston, S.; Lee, I., M. D.; Lewis, R.; McCarren, P.; Muncipinto, G.; Norinder, U.; Perry, M.; Duvall, J. R.; Kihlberg, J. Structural and conformational determinants of macrocycle cell permeability. *Nature Chem. Biol.* **2016,** *12*, 1065-1074.
27. Over, B.; McCarren, P.; Artursson, P.; Foley, M.; Giordanetto, F.; Grönberg, G.; Hilgendorf, C.; Lee, M. D.; Matsson, P.; Muncipinto, G.; Pellisson, M.; Perry, M. W. D.; Svensson, R.; Duvall, J. D.; Kihlberg, J. Impact of stereospecific intramolecular hydrogen bonding on cell permeability and physicochemical properties. *J. Med. Chem.* **2014,** *57*, 2746−2754.
28. Matsson, P.; Doak, B. C.; Over, B.; Kihlberg, J. Cell permeability beyond the rule of 5. *Adv. Drug Delivery Rev.* **2016,** *101*, 42-61.
29. Alex, A.; Millan, D. S.; Perez, M.; Wakenhut, F.; Whitlock, G. A. Intramolecular hydrogen bonding to improve membrane permeability and absorption in beyond rule of five chemical space. *Med. Chem. Commun.* **2011,** *2*, 669-674.
30. Whitty, A.; Zhong, M.; Viarengo, L.; Beglov, D.; Hall, D. R.; Vajda, S. Quantifying the chameleonic properties of macrocycles and other high-molecular-weight drugs. *Drug Discov. Today* **2016,** *21*, 712-717.
31. Rezai, T.; Yu, B.; Millhauser, G. L.; Jacobson, M. P.; Lokey, R. S. Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. *J Am Chem Soc* **2006,** *128*, 2510-1.
32. Matsson, P.; Kihlberg, J. How big is too big for cell permeability? *J. Med. Chem.* **2017,** *60*, 1662-1664.
33. Caron, G.; Ermondi, G. Updating molecular properties during early drug discovery. *Drug Discov. Today* **2017,** *22*, 835-840.
34. Kuhn, B.; Mohr, P.; Stahl, M. Intramolecular hydrogen bonding in medicinal chemistry. *J. Med. Chem.* **2010,** *53*, 2601-2611.

35. Desai, P. V.; Raub, T. J.; Blanco, M.-J. How hydrogen bonds impact P-glycoprotein transport and permeability. *Bioorg. Med. Chem. Lett.* **2012,** *22*, 6540–6548.

36. Rafi, S. B.; Hearn, B. R.; Vedantham, P.; Jacobson, M. P.; Renslo, A. R. Predicting and improving the membrane permeability of peptidic small molecules. *J. Med. Chem.* **2011,** *55*, 3163-3169.

37. Palm, K.; Stenberg, P.; Luthman, K.; Artursson, P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* **1997,** *14*, 568-571.

38. Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002,** *45*, 2615-2623.

39. Caron, G.; Ermondi, G. Molecular descriptors for polarity: The need of going beyond polar surface area. *Future Med. Chem.* **2016,** *8*, 2013-2016.

40. Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A. C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z. T.; Han, B.; Zhou, Y.; Wishart, D. S. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **2014,** *42*, D1091-7.

41. Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000,** *43*, 3714-3717.

42. Guimarães, C. R. W.; Mathiowetz, A. M.; Shalaeva, M.; Goetz, G.; Liras, S. Use of 3D properties to characterize beyond rule-of-5 property space for passive permeation *J. Chem. Inf. Model.* **2012,** *52*, 882-890.

43. Watts, K. S.; Dalal, P.; Tebben, A. J.; Cheney, D. L.; Shelley, J. C. Macrocycle Conformational Sampling with MacroModel. *J. Chem. Inf. Model.* **2014,** *54*, 2680-2696.

44. Chen, I. J.; Foloppe, N. Tackling the conformational sampling of larger flexible compounds and macrocycles in pharmacology and drug discovery. *Bioorg. Med. Chem.* **2013,** *21*, 7898-920.

45. Sondergaard, C. R.; Garrett, A. E.; Carstensen, T.; Pollastri, G.; Nielsen, J. E. Structural artifacts in protein-ligand X-ray structures: implications for the development of docking scoring functions. *J. Med. Chem.* **2009,** *52*, 5673-84.

46. Liebeschuetz, J.; Hennemann, J.; Olsson, T.; Groom, C. R. The good, the bad and the twisted: a survey of ligand geometry in protein crystal structures. *J. Comput. Aided Mol. Des.* **2012,** *26*, 169-83.

47. Holmstock, N.; Annaert, P.; Augustijns, P. Boosting of HIV protease inhibitors by ritonavir in the intestine: the relative role of cytochrome P450 and P-glycoprotein inhibition based on Caco-2 monolayers versus in situ intestinal perfusion in mice. *Drug Metab Dispos* **2012,** *40*, 1473-7.

48. Horie, K.; Tang, F.; Borchardt, R. T. Isolation and characterization of Caco-2 subclones expressing high levels of multidrug resistance protein efflux transporter. *Pharm Res* **2003,** *20*, 161-8.

49. Lazorova, L.; Hubatsch, I.; Ekegren, J. K.; Gising, J.; Nakai, D.; Zaki, N. M.; Bergstrom, C. A.; Norinder, U.; Larhed, M.; Artursson, P. Structural features determining the intestinal epithelial permeability and efflux of novel HIV-1 protease inhibitors. *J Pharm Sci* **2011,** *100*, 3763-72.

50. Li, Y.; Zhou, J.; Ramsden, D.; Taub, M. E.; O'Brien, D.; Xu, J.; Busacca, C. A.; Gonnella, N.; Tweedie, D. J. Enzyme-transporter interplay in the formation and clearance of abundant metabolites of faldaprevir found in excreta but not in circulation. *Drug Metab Dispos* **2014,** *42*, 384-93.

51. Lin, X.; Skolnik, S.; Chen, X.; Wang, J. Attenuation of intestinal absorption by major efflux transporters: quantitative tools and strategies using a Caco-2 model. *Drug Metab Dispos* **2011,** *39*, 265-74.

52. Uchida, M.; Fukazawa, T.; Yamazaki, Y.; Hashimoto, H.; Miyamoto, Y. A modified fast (4 day) 96-well plate Caco-2 permeability assay. *J Pharmacol Toxicol Methods* **2009,** *59*, 39-43.

53. White, P. W.; Llinas-Brunet, M.; Amad, M.; Bethell, R. C.; Bolger, G.; Cordingley, M. G.; Duan, J.; Garneau, M.; Lagace, L.; Thibeault, D.; Kukolj, G. Preclinical characterization of BI 201335, a C-terminal carboxylic acid inhibitor of the hepatitis C virus NS3-NS4A protease. *Antimicrob Agents Chemother* **2010,** *54*, 4611-8.

54. Witek, J.; Keller, B. G.; Blatter, M.; Meissner, A.; Wagner, T.; Riniker, S. Kinetic models of cyclosporin A in polar and apolar environments reveal multiple congruent conformational states. *J. Chem. Inf. Model.* **2016,** *56*, 1547−1562.

55. Leung, S. S.; Sindhikara, D.; Jacobson, M. P. Simple Predictive Models of Passive Membrane Permeability Incorporating Size-Dependent Membrane-Water Partition. *J Chem Inf Model* **2016,** *56*, 924-9.
56. Xiang, T. X.; Anderson, B. D. The relationship between permeant size and permeability in lipid bilayer membranes. *J Membr Biol* **1994,** *140*, 111-22.
57. Bissantz, C.; Kuhn, B.; Stahl, M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **2010,** *53*, 5061-5084.
58. Caron, G.; Vallaro, M.; Ermondi, G. High throughput methods to measure the propensity of compounds to form intramolecular hydrogen bonding. *Med. Chem. Commun.* **2017,** *8*, 1143–1151.
59. White, P. W.; Llinas-Brunet, M.; Amad, M.; Bethell, R. C.; Bolger, G.; Cordingley, M. G.; Duan, J.; Garneau, M.; Lagace, L.; Thibeault, D.; Kukolj, G. Preclinical characterization of BI 201335, a C-terminal carboxylic acid inhibitor of the hepatitis C virus NS3-NS4A protease. *Antimicrob. Agents Chemother.* **2010,** *54*, 4611-4618.
60. Artursson, P.; Palm, K.; Luthman, K. Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv. Drug Delivery Rev.* **2001,** *46*, 27-43.
61. Hosseinzadeh, P.; Bhardwaj, G.; Khipple Mulligan, V. K.; Shortridge, M. D.; Craven, T. W.; Pardo-Avila, F.; Rettie, S. A.; Kim, D. E.; Silva, D.-A.; Ibrahim, Y. M.; Webb, I. K.; Cort, J. R.; Adkins, J. N.; Varani, G.; Baker, D. Comprehensive computational design of ordered peptide macrocycles. *Science* **2017,** *358*, 1461–1466.
62. Shalaeva, M.; Caron, G.; Abramov, Y. A.; O'Connell, T. N.; Plummer, M. S.; Yalamanchi, G.; Farley, K. A.; Goetz, G. H.; Philippe, L.; Shapiro, M. J. Integrating intramolecular hydrogen bonding (IMHB) considerations in drug discovery using ΔlogP as a tool. *J. Med. Chem.* **2013,** *56*, 4870-4879.
63. Rezai, T.; Bock, J. E.; Zhou, M. V.; Kalyanaraman, C.; Lokey, R. S.; Jacobsen, M. P. Conformational Flexibility, Internal Hydrogen Bonding, and Passive Membrane Permeability: Successful in Silico Prediction of the Relative Permeabilities of Cyclic Peptides. *J. Am. Chem. Soc.* **2006,** *128*, 14073-14080.
64. Hubatsch, I.; Ragnarsson, E. G.; Artursson, P. Determination of drug permeability and prediction of drug absorption in Caco-2 monolayers. *Nat. Protoc.* **2007,** *2*, 2111-2119.

65. Low, Y. W.; Blasco, F.; Vachaspati, P. Optimised method to estimate octanol water distribution coefficient (logD) in a high throughput format. *Eur. J. Pharm. Sci.* **2016,** *92*, 110-6.

66. Schonherr, D.; Wollatz, U.; Haznar-Garbacz, D.; Hanke, U.; Box, K. J.; Taylor, R.; Ruiz, R.; Beato, S.; Becker, D.; Weitschies, W. Characterisation of selected active agents regarding pKa values, solubility concentrations and pH profiles by SiriusT3. *Eur. J. Pharm. Biopharm.* **2015,** *92*, 155-70.

# TABLE OF CONTENTS GRAPHIC