

Structural and conformational determinants of macrocycle cell permeability

Björn Over¹, Pär Matsson^{2,3}, Christian Tyrchan⁴, Per Artursson^{2,3}, Bradley C Doak⁵, Michael A Foley^{6,9}, Constanze Hilgendorf⁷, Stephen E Johnston⁶, Maurice D Lee IV^{6,9}, Richard J Lewis⁴, Patrick McCarren⁶, Giovanni Muncipinto^{6,9}, Ulf Norinder⁸, Matthew W D Perry⁴, Jeremy R Duvall^{6,9*} & Jan Kihlberg^{5*}

Macrocycles are of increasing interest as chemical probes and drugs for intractable targets like protein–protein interactions, but the determinants of their cell permeability and oral absorption are poorly understood. To enable rational design of cell-permeable macrocycles, we generated an extensive data set under consistent experimental conditions for more than 200 non-peptidic, *de novo*-designed macrocycles from the Broad Institute's diversity-oriented screening collection. This revealed how specific functional groups, substituents and molecular properties impact cell permeability. Analysis of energy-minimized structures for stereo- and regioisomeric sets provided fundamental insight into how dynamic, intramolecular interactions in the 3D conformations of macrocycles may be linked to physicochemical properties and permeability. Combined use of quantitative structure–permeability modeling and the procedure for conformational analysis now, for the first time, provides chemists with a rational approach to design cell-permeable non-peptidic macrocycles with potential for oral absorption.

Compounds with properties outside Lipinski's rule of 5 (Ro5) (ref. 1) are attractive in drug discovery and chemical biology, as they can modulate novel targets that have difficult, large, and featureless binding sites^{2–5}. Furthermore, improved selectivity, which may reduce off-target toxicity, has been reported when molecular size and complexity increase^{4,6}. Macrocycles have raised particular interest^{4,6–8}, and a survey of drugs and clinical candidates found that they are enriched in orally bioavailable chemical space beyond the Ro5 (bRo5) (ref. 3). Prominent classes of macrocyclic drugs in bRo5 space include broad-spectrum antibacterials, immunosuppressants and hepatitis C virus (HCV) protease inhibitors. Recent investigations suggest that the enrichment of macrocycles in bRo5 drug space is a result of an increased ability to bind to 'difficult to drug', flat- and groove-shaped binding sites with high specificity^{2,8}, coupled with improved cell permeability, metabolic stability and oral bioavailability compared to linear analogs^{3,6,7}.

Use of both macrocyclic and non-macrocyclic compounds that reside in chemical space at the borders of or beyond the Ro5 to modulate difficult targets is not without potential risks³. Increased molecular weight has been associated with reduced cell permeability and solubility, as well as increased metabolic liabilities, all of which reduce bioavailability. Some of these difficulties are illustrated by macrocyclic peptides, which typically have limited cell permeability due to their high number of polar functionalities, and also sometimes suffer from low solubility. Several recent studies describe approaches to improve cell permeability for cyclic peptides, resulting in moderate oral bioavailabilities in rodents^{9–12}. These developments constitute an important achievement, but the design of such peptides is currently restricted to chemical space where

polarity is shielded through N-methylation, alteration of stereochemistry and/or use of large and hydrophobic side chains. In line with this, cyclosporin A is the only example of an orally administered compound among the 30 approved macrocyclic peptide drugs⁷. For drug targets and indications where cell permeability and oral administration are required (for example, HCV protease inhibitors)^{3,13}, use of less peptidic, *de novo*-designed and natural-product-like macrocycles have been found to be advantageous. Consequently, the number of *de novo*-designed oral macrocycles in clinical trials is increasing, even though they are often synthetically more challenging than cyclic peptides⁷.

Insight into the molecular determinants of macrocycle cell permeability is very limited and is based on scattered data for a small number of drugs and clinical candidates^{3,7}, some natural products⁸ and cyclic peptide model systems^{9–12,14–16}. In order to identify molecular features that provide drug-like physicochemical properties and cell permeability in *de novo*-designed macrocycles at the border of and in beyond Ro5 space, we profiled >200 structurally diverse macrocycles from the Broad Institute's diversity-oriented synthesis (DOS) screening collection¹⁷. This collection is inspired by non-peptidic natural products¹⁸ and consists of complete matrices of enantiomers and stereoisomers. It thereby provides a unique opportunity to understand molecular properties that control bRo5 and macrocycle cell permeability, while at the same time, systematically dissecting the influence of structural motifs and stereochemistry. We were able to identify specific functional groups, molecular properties and conformational aspects that impact cell permeability. Translation of our findings allowed for the development of a two-tiered approach to designing macrocycles with improved cell permeability.

¹Cardiovascular and Metabolic Diseases, Innovative Medicines and Early Development Biotech Unit, AstraZeneca R&D Gothenburg, Mölndal, Sweden.

²Department of Pharmacy, Uppsala University, Uppsala, Sweden. ³Uppsala University Drug Optimization and Pharmaceutical Profiling Platform (UDOPP), a node at the Chemical Biology Consortium Sweden, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ⁴Respiratory, Inflammation and Autoimmunity Diseases, Innovative Medicines and Early Development Biotech Unit, AstraZeneca R&D Gothenburg, Mölndal, Sweden. ⁵Department of Chemistry, Uppsala University, Uppsala, Sweden. ⁶Center for the Science of Therapeutics, Broad Institute, Cambridge, Massachusetts, USA. ⁷Drug Safety and Metabolism, Safety & ADME Translational Sciences, AstraZeneca R&D Gothenburg, Mölndal, Sweden. ⁸Swedish Toxicology Sciences Research Center, Södertälje, Sweden. ⁹Present addresses: Tri-Institutional Therapeutics Discovery Institute, New York, New York, USA (M.A.F.) and Ensemble Therapeutics, Cambridge, Massachusetts, USA (M.D.L., G.M. and J.R.D.). *e-mail: jduvall@ensembletx.com or jan.kihlberg@kemi.uu.se

RESULTS

A representative selection of DOS macrocycles

In order to profile macrocycles at the border of and beyond the Ro5, we first filtered the Broad Institute's DOS collection for compounds with ring sizes of 12 or more atoms and molecular weight (MW) above 400 Da. This resulted in a subset of 20,998 macrocycles. While the majority of the overall collection (>100,000 compounds) conforms to the Ro5 guidelines for oral bioavailability, 70% of the macrocyclic subset fall outside the Ro5 boundaries in at least one molecular property, and 49% break two or more Ro5 guidelines. The frequency of Ro5 violations by property decreases in the order of MW > hydrogen bond acceptors (HBA) > lipophilicity (logP) > hydrogen bond donors (HBD) (Fig. 1a, bottom). Importantly, the DOS macrocycles have MW of 400–800 Da, the same range as *de novo*-designed oral macrocycles currently in clinical trials⁷.

The selected DOS macrocycles were composed of three main chemical scaffolds synthesized through Huisgen cycloaddition (click

chemistry; 1,591 compounds; Fig. 1b, 1), head-to-tail macrocyclization (head-to-tail; 7,054 compounds; 2) or ring-closing metathesis (RCM; 11,896 compounds; 3) (refs. 19,20); plus a smaller set composed of miscellaneous scaffolds (4, 5) (ref. 21). The miscellaneous set comprised nine types of macrocycles, including 12 different ring assemblies, in 457 compounds. To reflect this diversity, 46 representative compounds were chosen by visual inspection, always including matched pairs of regio- or stereoisomers. For the three large sublibraries, a cheminformatics approach was used to ensure a representative coverage of chemical property space and stereochemical diversity. Compounds were connected in a chemical network (Fig. 1a) based on their pairwise similarity in key molecular properties; MW, molecular volume, calculated lipophilicity (AlogP), the surface areas of polar and non-polar atoms, and HBAs and HBDs. Representative compounds from each structural sublibrary were subsequently selected by picking the medoids from compound clusters across the entire similarity network, along with stereo- and/or

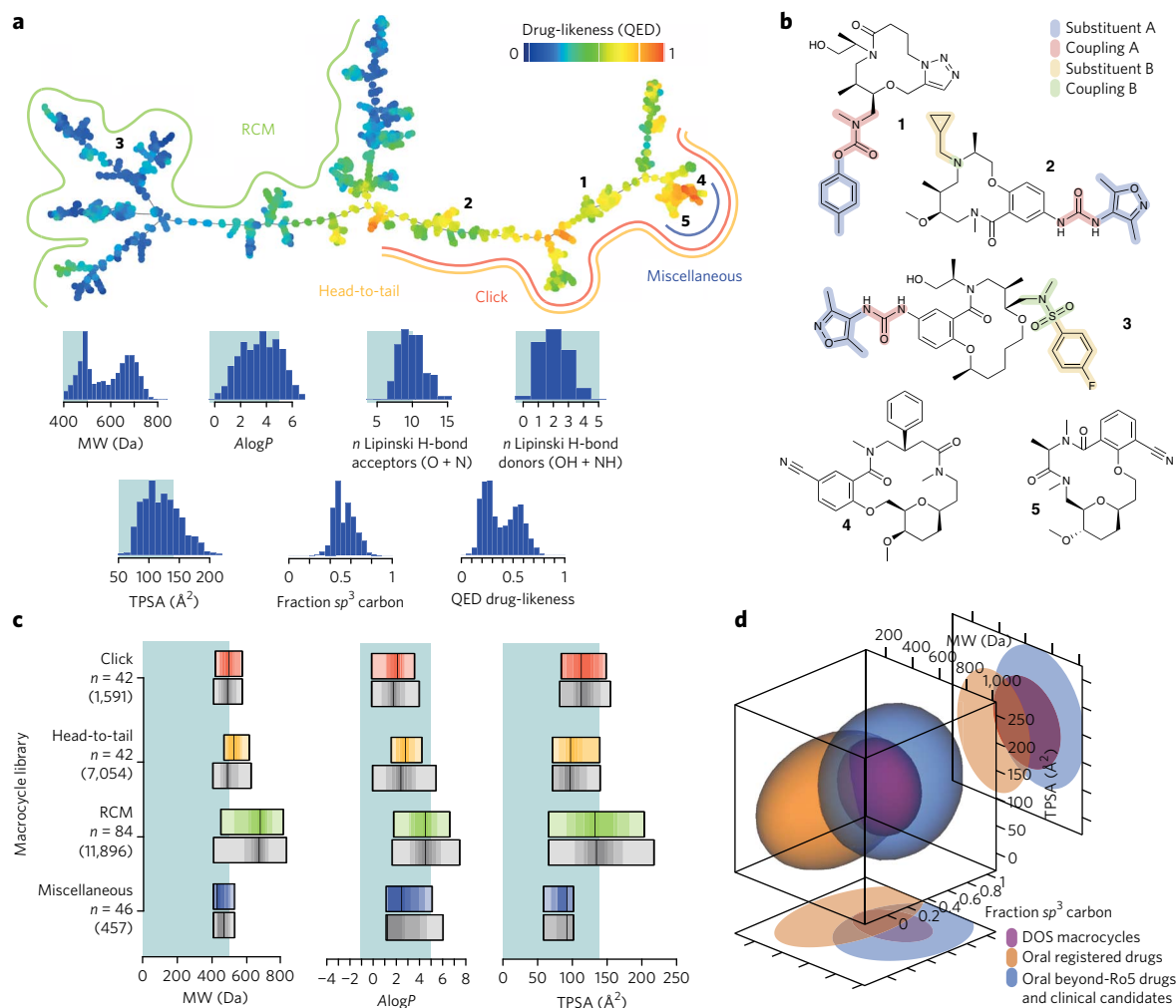


Figure 1 | Overview of properties of the representative macrocycle selection. (a) Top, chemical similarity network of Broad Institute DOS macrocycles with MW > 400 Da ($n = 20,998$). Compounds are represented by points, which are connected based on their molecular property similarity and colored by the quantitative estimate of drug likeness (QED)⁴¹. Labels indicate the predominant sublibrary in different regions of the network, and numbers show the location of compounds 1–5. Molecular property distributions of the 20,998 macrocycles are shown below the network, with shading indicating Ro5-compliant property ranges and the PSA cutoff for oral absorption. (b) Examples of structures from the macrocycle series prepared by click chemistry (Click; 1), head-to-tail chemistry (Head-to-tail; 2), ring-closing metathesis (RCM; 3), and a diverse set of compounds with varying core ring systems (Miscellaneous; 4 and 5). Substituents and coupling groups are indicated in color. (c) Molecular properties of macrocycles selected for profiling. Properties of the selected compounds (top rows) are compared to those of all macrocyclic compounds in the same structural series (bottom rows, gray, numbers in parentheses). Vertical lines indicate median values, and increasingly darker shading indicates the 0–100th, 10–90th, 20–80th, 30–70th, and 40–60th percentile ranges. (d) The chemical space of DOS macrocycles (purple) in relation to that of registered oral drugs (orange) and oral bRo5 drugs and clinical candidates (blue). Ellipsoids show 10–90th percentile property ranges.

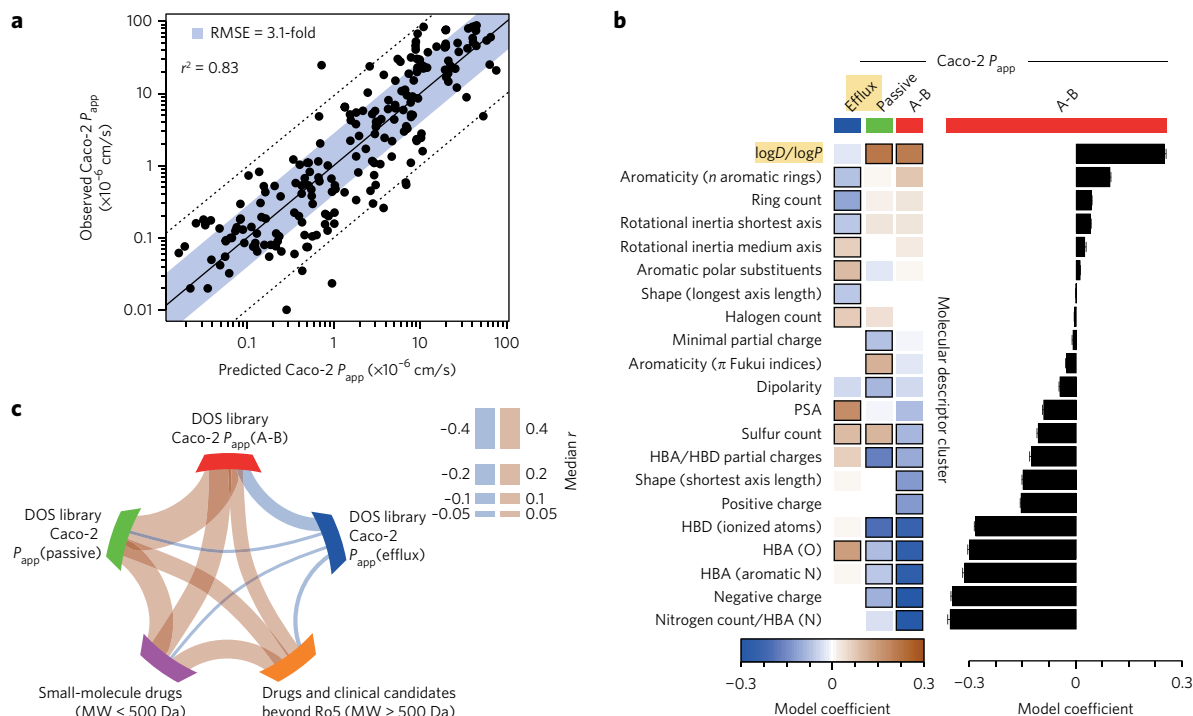


Figure 2 | Molecular properties influencing cell permeability of macrocycles. (a) External predictions of the PLS-derived structure–permeability model. The unity line and ± 10 fold error range are shown as solid and dashed lines, respectively. The shaded region indicates the external root mean squared error of prediction (RMSE). (b) Molecular properties influencing macrocycle cell permeability and efflux. Molecular descriptors are divided into molecular property groups (left) based on their intercorrelation in the set of examined macrocycles. The ten most influential descriptor groups in models of macrocycle apical-to-basolateral permeability (red, also visualized in the waterfall plot, far right), passive permeability (green), and transporter-mediated efflux rates (blue) are indicated with solid borders (middle). Brown color indicates a positive impact on the permeability endpoint and blue color a negative impact. Darker shades indicate a greater influence. Error bars represent s.d. (c) Average pairwise similarity between models trained on macrocycle permeability (P_{app}), passive permeability (green) and efflux (blue) for the DOS macrocycles, Caco-2 cell permeability of oral small-molecule drugs (MW < 500 Da, purple) and on the administration route of beyond Ro5 drugs and clinical candidates (MW > 500 Da, oral versus non-oral administration; orange). Model similarity is calculated as the median correlation between model coefficients. Center brown colors indicate a positive correlation, blue a negative one, with the magnitude of correlation reflected in the width of the line. Model predictions and coefficients are the mean of 100 models, each trained on a randomized subset of the full data set.

regioisomers thereof (Supplementary Results, Supplementary Table 1). Overall, 214 macrocycles were selected that were representative of the macrocycles in the Broad DOS collection (Fig. 1c) and that have similar physicochemical properties to *de novo*-designed macrocyclic clinical candidates (Fig. 1d)⁷. Significant proportions (71% and 34%, respectively) of the selected macrocycles fall outside the Ro5 boundaries in at least one, or in two or more, molecular properties.

Characterization of the macrocycles

We measured solubility, lipophilicity ($\log D_{7.4}$), pK_a (when relevant) and permeability across Caco-2 cell monolayers in the apical-to-basolateral (P_{app} (A-B)) and basolateral-to-apical (P_{app} (B-A)) directions for all 214 selected macrocycles (Supplementary Tables 2–10; Supplementary Data Sets 1 and 2). The majority (78%) of the macrocycles had high solubility (>50 μ M), and although solubility began to decrease at $\log D > 3$, less than 10% had a solubility <10 μ M (Supplementary Fig. 1). The measured $\log D_{7.4}$ values ranged from –1.6 to 5.3 (median 1.8), which are the detection limits of the assay. Approximately 45% of the macrocycles had drug-like $\log D$ values²² between 1 and 3, and an additional 29% had a $\log D > 3$. The Caco-2 cell permeability (P_{app} (A-B)) ranged from undetectable (<0.01; $n = 14$; these compounds were excluded from further analyses because of poor recovery and/or undetectable permeation) to 88×10^{-6} cm/s (median 1.15×10^{-6} cm/s), with 54% showing a

moderate-to-high cell permeability (P_{app} (A-B) > 1×10^{-6} cm/s) that is predictive of medium-to-high absorption from the intestine^{23–25}. More than 75% of all compounds exhibited a significant transporter-mediated cellular efflux, as revealed by efflux ratios (P_{app} (B-A)/ P_{app} (A-B)) > 2 (Supplementary Tables 2–10), with the most pronounced effects observed for compounds with low-to-medium cell permeability (P_{app} (A-B) < 5×10^{-6} cm/s; Supplementary Fig. 1). Ratios significantly smaller than that of the passive permeability control metoprolol were not observed, suggesting that efflux was the dominating carrier-mediated process for these macrocycles. Efflux was therefore investigated further for a subset of the macrocycles representing the stereo- or regioisomers with the largest variations in P_{app} (A-B) (2- to 25-fold) and measured $\log D_{7.4}$ (up to 2 log units), and including ten additional compounds that ensured representative coverage of chemical space (Supplementary Fig. 2; Supplementary Tables 11–32). For this subset, Caco-2 cell permeability (P_{app} (A-B)/ P_{app} (B-A)) was also measured in the presence of an inhibitor cocktail (quinidine, sulfasalazine and benzbromarone; Online Methods) 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*) (refs. 26,27). All efflux ratios were reduced by 90–100% when the inhibitor cocktail was added (Supplementary Fig. 3), allowing us to study intrinsic cell permeability and efflux transporter effects separately.

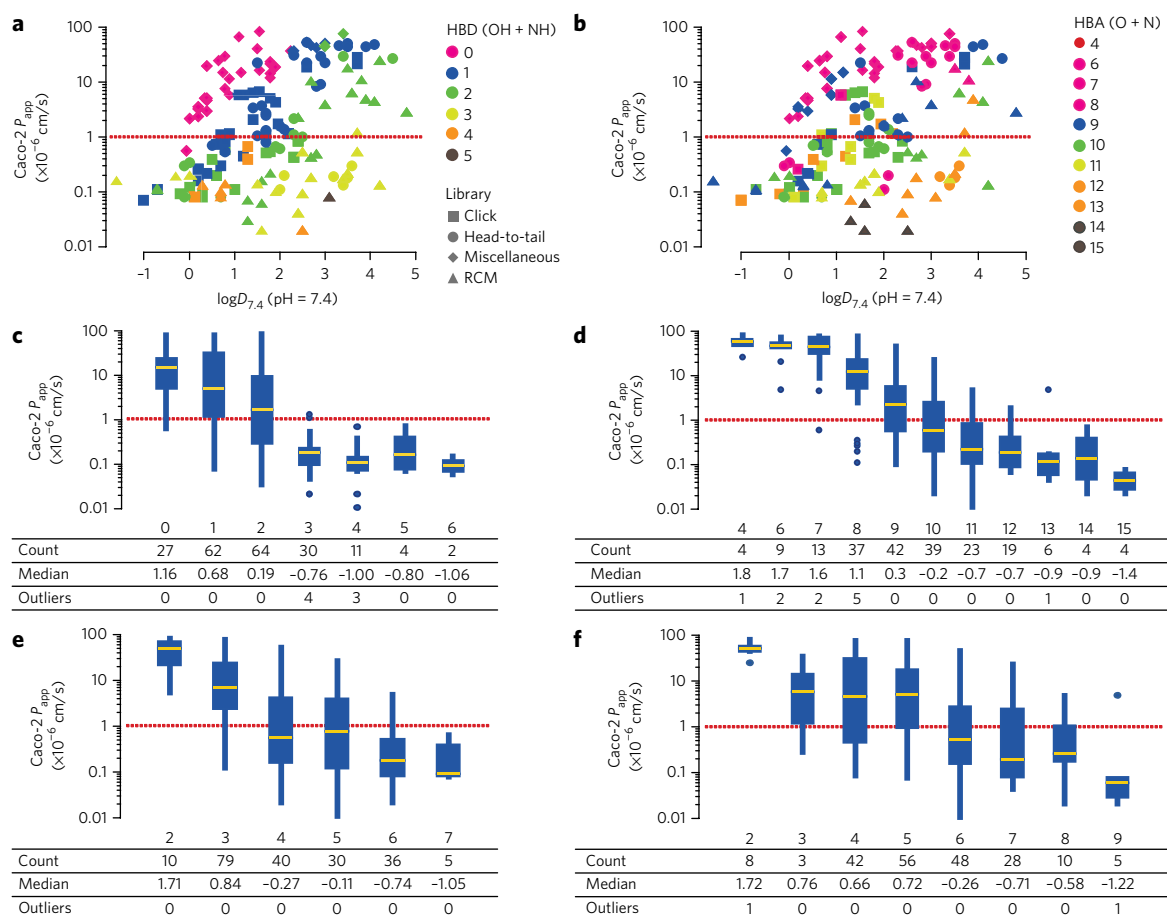


Figure 3 | Influence of lipophilicity and polar features on cell permeability. (a,b) Measured $\log D_{7.4}$ and Caco-2 cell permeability for the macrocycle screening set ($n = 200$), colored by number of hydrogen bond donors (HBD; a) or acceptors (HBA; b) and shaped by chemical series. (c–f) Cell permeability (Caco-2 P_{app} (A–B)) as a function of increasing HBD (c), HBA (d), nitrogen atom (e), and oxygen atom (f) count, with medians shown as horizontal lines. Compounds having a $P_{\text{app}} < 1 \times 10^{-6} \text{ cm/s}$ (dotted red lines) are likely to have low absorption on oral administration.

Molecular properties defining cell permeable macrocycles

Structure–permeability relationships correlating macrocycle cell permeability to calculated descriptors were analyzed using three different techniques (Supplementary Data Set 3). This resulted in models having similar predictive power, as revealed by the statistics for linear (partial least-squares projection (PLS): external prediction $r^2 = 0.83$; external root mean squared error of prediction (RMSE) = 0.50), nonlinear (support vector machine (SVM): external $r^2 = 0.84$ and RMSE = 0.41) and random forest (RF: external $r^2 = 0.81$ and RMSE = 0.45) regression models. Since the impact of specific molecular properties is more readily assessed from PLS models, further analyses were based on that approach (Fig. 2; Supplementary Figs. 4 and 5).

It has been suggested that conventional metrics that are used for characterization of classical oral small-molecule drugs cannot be applied to macrocycles^{8,28}. However, the PLS-derived quantitative structure–permeability (QSPR) models (Fig. 2a,b) identified lipophilicity (calculated $\log D_{7.4}$) as the most important descriptor for predicting increased cell permeability in the 214 DOS macrocycles, whereas hydrogen bonding descriptors (in particular, nitrogen-based functionalities), molecular charge and polarity were the most influential permeability limiting factors (Fig. 2b; Supplementary Fig. 6). Similar trends were seen for models of efflux-inhibited macrocycle cell permeability (P_{app} (Passive)) (Fig. 2b,c). The molecular descriptors included in the final optimized permeability models showed good overlap with both a reference set of small-molecule drugs (MW < 500 Da)²⁹ and a set of ‘beyond-Ro5 drugs and clinical

candidates’ (MW > 500 Da) (Fig. 2c; Supplementary Tables 33–38) that were included to allow comparison with compounds at the outer borders of cell permeability and oral bioavailability³.

Regardless of modeling technique, predictivity improved when experimentally determined $\log D_{7.4}$ was incorporated in the models, most likely because this parameter captures stereoisomer-specific membrane interactions that are not described in 2D calculations (PLS: external $r^2 = 0.93$ and RMSE = 0.36; SVM: $r^2 = 0.87$ and RMSE = 0.36; RF: $r^2 = 0.82$ and RMSE = 0.43). Despite this importance of lipophilicity, a univariate correlation between experimentally measured $\log D_{7.4}$ and cell permeability was not obvious at first glance; however, stratifying the data by the number of HBDs in each macrocycle revealed a clear trend (Fig. 3a). For a given $\log D$, the cell permeability decreases with an increasing number of HBDs, and correspondingly, $\log D$ must be increased to retain cell permeability with an increasing number of HBDs. A similar correlation between the total number of HBAs and $\log D$ was not as apparent (Fig. 3b). Notably, compounds with three or more HBDs show low cell permeability (P_{app} (A–B) < $1 \times 10^{-6} \text{ cm/s}$) and often high efflux ratios (Fig. 3c; Supplementary Figs. 7–9). A continuous permeability decrease was observed for compounds with >7 HBA, and the median cell permeability (< $1 \times 10^{-6} \text{ cm/s}$) was low for compounds with 10 or more HBAs (Fig. 3d). The molecular-property-based permeability models suggest a stronger impact of nitrogen hydrogen bond acceptors over oxygen hydrogen bond acceptors; this was also supported by a negative correlation between cell permeability and the total number of nitrogen atoms (Fig. 3e,f).

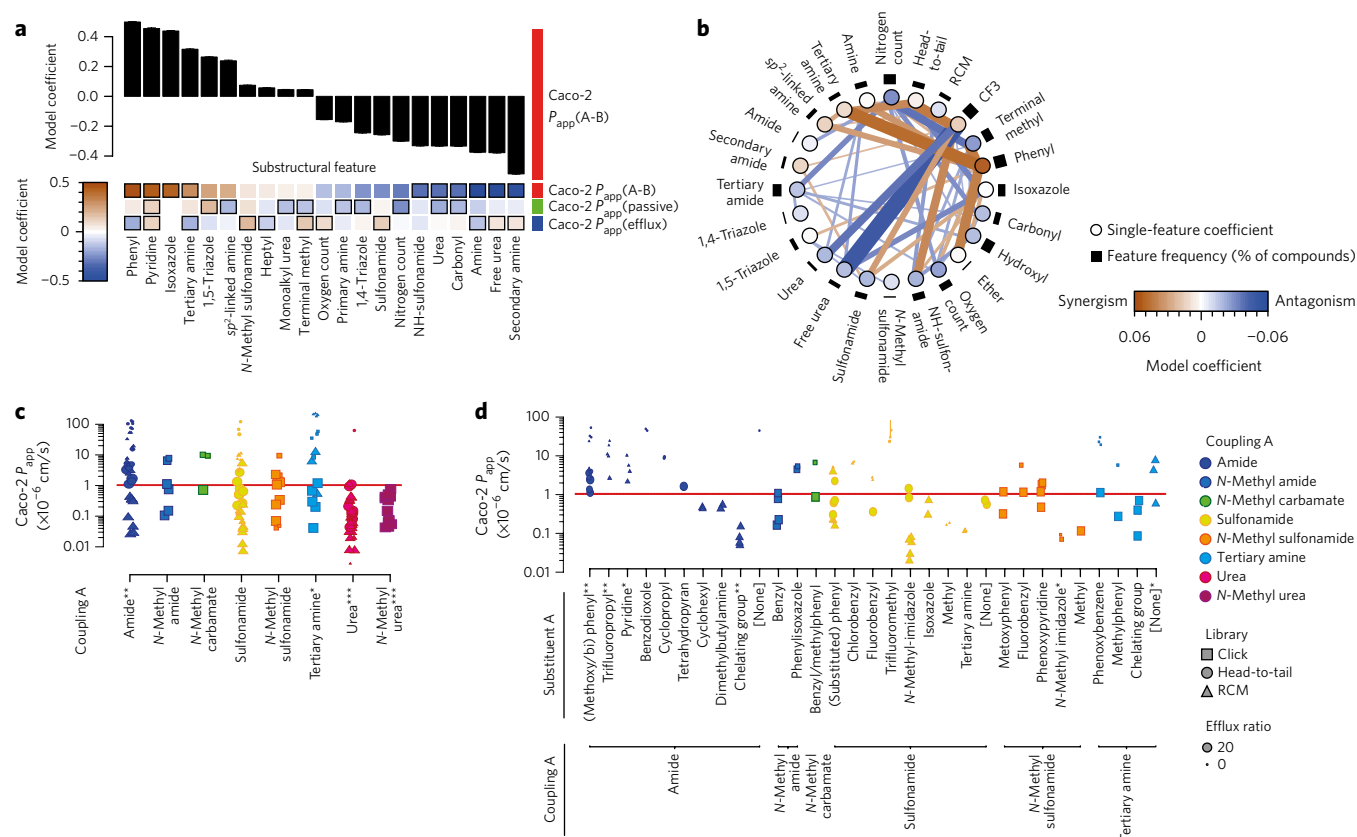


Figure 4 | Influence of substructural features on cell permeability. (a) Impact of individual substructural features on cell permeability and efflux. Models were trained using individual substructural features as descriptors, excluding pairwise interaction terms. Brown color indicates a positive impact on the permeability endpoint, and blue color a negative impact. Darker shades indicate a greater influence. Model predictions and coefficients are the mean of 100 models, each trained on a randomized subset of the full data set. Error bars represent s.d. (b) Top pairwise interactions between substructures, indicating combinatorial effects on macrocycle apical-to-basolateral cell permeability. Brown color indicates feature synergism, and blue color feature antagonism. Circle color shows the average influence of the individual features. The heights of the bars indicate the prevalence of the feature in the studied compounds. (c) Dependence of Caco-2 cell permeability on the nature of coupling group A (red shadow in Fig. 1b). Symbols as in d. (d) Combined influence of coupling group A and substituent A (blue shadow in Fig. 1b) on cell permeability for the three main chemical scaffolds, shaped by macrocycle chemical series and sized by efflux ratio. The ‘chelating group’ is an *ortho*-amino anilide. The red line indicates the cutoff between low and moderate-to-high cell permeability. Substructures significantly associated with permeability above or below this cutoff are indicated * $P < 0.05$, ** $P < 0.01$ or *** $P < 0.001$, as determined using Student’s *t*-test. Significance is shown only for groups of four or more compounds.

Quantitative models of efflux were dominated by other descriptors (Fig. 2b,c). Efflux was primarily correlated to high numbers of oxygen hydrogen bond acceptors, high PSA and the presence of sulfur (Fig. 2b). In our compound set, the latter reflects the presence of sulfonamides, a well-known motif for substrates of BCRP/ABCG2 and P-gp/ABCB1 (refs. 30,31). Furthermore, descriptors of molecular shape and size correlated with efflux, with larger cross-sectional area generally associated with higher efflux. The considerable conformational dependence of shape suggests that this likely reflects a general size increase rather than a shape-specific effect (Supplementary Figs. 10–12). Conversely, an increased number of rings, particularly aromatic rings such as (substituted) phenyls, was the dominant efflux-limiting feature.

Substructure–cell permeability relationships

The systematic design of the DOS macrocycle collection enabled us to investigate the influences of structural features on cell permeability using a substructure structure–activity relationship (SAR) approach. We considered variation in the functional groups that link scaffolds to substituents (‘coupling groups’), the individual substituents, and stereo- and regiochemistry (Fig. 1b; Supplementary Figs. 13–23). First, we investigated the general influences of structural features using the same multivariate regression methodology

as used for the quantitative molecular-property-based PLS models (Fig. 4a; full overview in Supplementary Fig. 6). This revealed that phenyl and pyridyl groups, isoxazoles and tertiary amines had a large positive impact on cell permeability, whereas ureas, carbonyl groups, secondary amines and sulfonamides were unfavorable. Interestingly, the 1,5-triazole click macrocycles always showed higher intrinsic cell permeabilities and generally lower efflux ratios than regioisomeric 1,4-triazoles with identical substituents (Supplementary Fig. 24; Supplementary Tables 12–16, 39 and 40). Random-forest modeling supported these findings (Supplementary Table 41; Supplementary Figs. 25 and 26); highlighting that phenyl, trifluoromethyl and pyridyl groups increase permeability, whereas it is reduced by high counts of oxygen and/or nitrogen atoms and incorporation of terminal methyl groups, ureas and/or hydroxyl groups.

Next, we expanded the PLS models with all pairwise combinations of structural features, with the aim of deducing combinatorial effects on uninhibited macrocycle permeability (Supplementary Figs. 27–29). Significant cross terms suggested synergistic or antagonistic effects between pairs of features. For example, phenyl groups and tertiary amines both increased cell permeability and showed a strong synergistic effect with each other (Fig. 4b). Additionally, both groups correlated with good cell permeability through synergistic

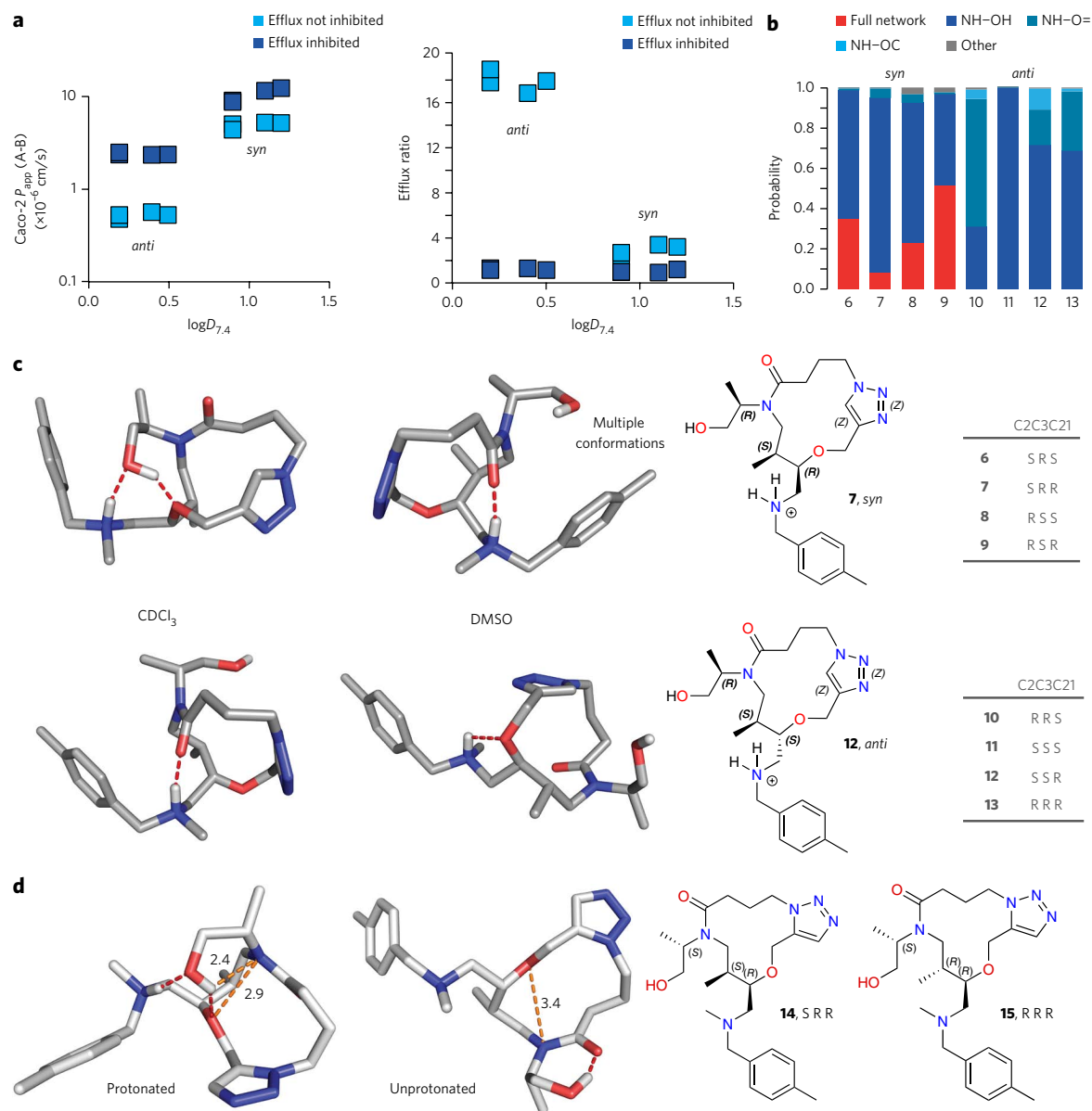


Figure 5 | Explanation of stereospecific differences in Caco-2 cell permeability and efflux for stereo- and regioisomeric macrocycles. (a) Cell permeability and efflux ratio versus lipophilicity ($\log D_{7.4}$) for **6-13**. Experiments were performed with (dark blue) and without (light blue) efflux transporter inhibitors. (b) Intramolecular polar interactions (red: full network $NH^+-OH-OC$ or $NH^+-OH-O=C$; dark blue: NH^+-OH ; petrol: $NH^+-O=C$; light blue: NH^+-OC ; gray: other) determined by a Boltzmann-weighted microstate probability analysis of all energy minimized states within 15 kcal/mol for *syn* (**6-9**) and *anti* stereoisomers (**10-13**). (c) Minimum energy conformations for the protonated form of **7** (top row) that were verified by NMR spectroscopy in $CDCl_3$ (left) and $DMSO-d_6$ (right). Conformations populated by protonated **12** (bottom row) in different ratios in $CDCl_3$ and $DMSO-d_6$. Intramolecular hydrogen bonds are indicated by red dotted lines. Tables (far right) show the stereochemistry for compounds **6-13**. (d) Minimum energy conformations for **14** in its protonated and unprotonated form. For the protonated form the marked distances are smaller (2.4 Å, 2.9 Å) than the atomic van der Waals radii, resulting in a steric clash.

cross terms to several other groups such as ethers and sp^2 -linked amines. Trifluoromethyl groups, which also contributed to permeability, showed positive cross terms with sulfonamides, suggesting use of this combination to overcome sulfonamide-induced efflux. Conversely, high oxygen and nitrogen counts individually had a negative impact on cell permeability. This effect was exacerbated when the two features were combined, and in addition, their cross terms with other groups were also antagonistic. Similarly, all top cross-term combinations that included ureas were negative, prohibiting high cell permeability. The dominant role of ureas—which is most likely a combined effect of decreased passive permeability and increased efflux (Fig. 4a)—was further illustrated by their antagonism

with trifluoromethyl groups, a typically positive contributor to permeability (Fig. 4b).

Collectively, our results indicate that the substituents and the coupling groups that connect them to the macrocyclic ring have a significant impact on the cell permeability of the macrocyclic cores, which all contain a limited number of polar groups, as well as HBD and HBAs. Further analysis of the coupling groups, and their combination with substituents, reiterated that ureas and *N*-methyl ureas consistently displayed very low cell permeabilities (Fig. 4c). The other coupling groups had more variable impacts on cell permeability depending on the attached substituent (Fig. 4d), and several coupling group–substituent combinations resulted in medium to

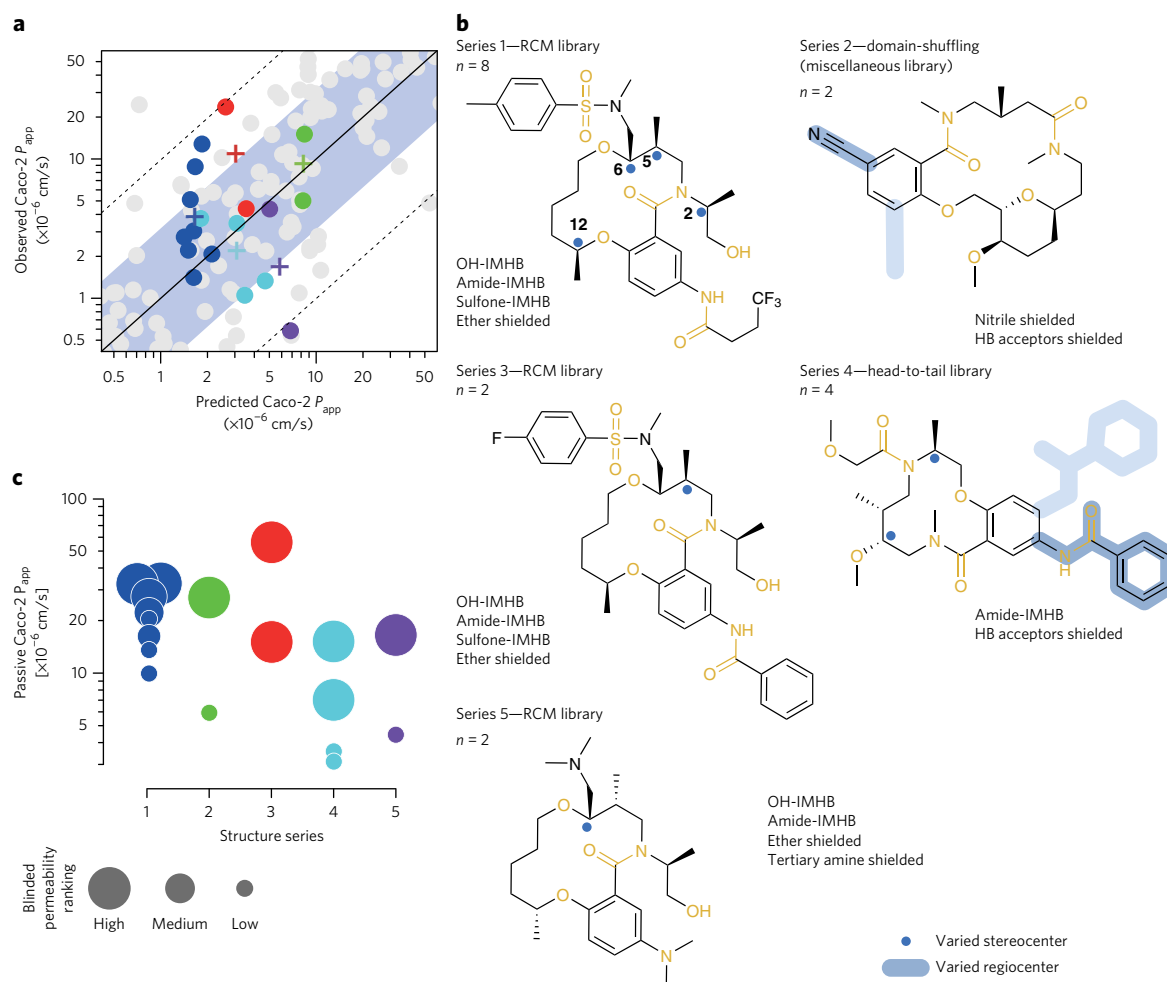


Figure 6 | Permeability predictions for *de novo*-designed macrocycles. (a) Observed, uninhibited Caco-2 cell permeability versus permeability predicted by quantitative structure–permeability modeling for five stereo- and regioisomeric series of macrocycles. The average observed and predicted permeability for each series is shown as a plus sign, with mean prediction errors being 2.1-fold. The unity line and ten-fold error range are shown as solid and dashed lines, respectively, and the average prediction error for the complete compound set is shown in blue. (b) Structural features used for scoring the relative permeability within each stereo- and regioisomeric series. The scored polar functionalities are highlighted in gold in the respective structures. The stereo- and/or regiocenters that were varied in each of the series are indicated with blue dots and shading, respectively. Full structures of all compounds are shown in **Supplementary Tables 18, 20, 22, 27 and 29**. (c) Blinded predictions of relative permeability within each isomeric series. Compounds within each series were ranked according to the criteria in **b**. Larger symbol size indicates a higher predicted permeability.

high cell permeability despite high efflux ratios. For example, macrocycles with amide or sulfonamide linkers can be cell permeable if they carry lipophilic substituents such as (halogenated) phenyl groups or trifluoromethyl groups. In contrast to ureas, sulfonamide linkers were primarily associated with increased efflux³¹; they were involved in nearly all top cross terms in PLS models trained on efflux data, but virtually absent in models of passive permeability (**Supplementary Figs. 30–34**). Analogous to the coupling groups, most substituents showed context-dependent effects on cell permeability; but even lipophilic substituents like fluorobenzyl groups did not salvage permeability for compounds with three or more HBDs (**Supplementary Fig. 8**). A few substituents, particularly a chelating *ortho*-amino anilide and *N*-methyl imidazoles, stood out by being consistently associated with low cell permeability (**Fig. 4d**).

Impact of regio- and stereochemistry on cell permeability

The influence of regio- and stereochemistry was clear, with nearly half of the macrocycles showing a 2- to 25-fold difference in P_{app} (A-B) between isomers (**Supplementary Tables 2–9**). This variability was linked to differences in both passive cellular permeability (measured in the presence of efflux inhibitors) and

transporter-mediated efflux rates, with up to 8- and 60-fold differences in the respective endpoint. In our compound selection we primarily aimed to maximize coverage of physicochemical and stereochemical diversity; therefore, enantiomeric pairs were only included for two clusters (click compounds **6–13**, **Fig. 5**, **Supplementary Table 39**; and RCM compounds **16–23**, **Supplementary Table 27**). While more than ten-fold differences in $\log D_{7.4}$ and permeability were observed between diastereoisomers in these clusters, enantiomers had similar $\log D_{7.4}$ values (differences <0.3 log units) and passive and uninhibited permeabilities (differences <1.4 -fold and <2.2 -fold, respectively). Flux ratios indicated modest enantiomer-selective efflux, with no signs of transporter-mediated uptake.

The often marked stereochemical differences observed in pharmacokinetic properties are rarely captured by predictions from 2D-QSPR models; the eight stereoisomers **6–13** constitute an illustrative example in that all were predicted to have a $\log D_{7.4}$ of 2.0 and P_{app} (A-B) in a narrow range between 2.6 and 3.4×10^{-6} cm/s. Experimentally, the set was split into two groups, which displayed significant differences in lipophilicity, passive cell permeability and efflux ratios, while having high (≥ 100 μ M) solubility (**Fig. 5a**; **Supplementary Table 39**). This division was correlated to the relative

configuration of the two stereocenters within the macrocyclic ring and investigations in the presence of selective efflux inhibitors indicated that P-gp^{26,27,32} was responsible for the differences in efflux ratios (Supplementary Table 40).

Investigating stereospecific cell permeability

To better understand the limitations of the 2D-QSPR models, we investigated the underlying structural features driving the differences in cell permeability between the *syn* and *anti* forms 6–9 and 10–13, respectively. As compounds 6–13 are protonated at physiological pH ($pK_a \sim 8.5$ for the *tert*-amine in 6–13), conformational ensembles of their charged forms were generated to understand the cell-permeability differences. The analysis indicated that the charged amine was shielded from solvent through an intramolecular hydrogen bond network with the hydroxyl group and the ether or the amide carbonyl group in up to 50% of the low-energy conformations ($\Delta E < 15$ kcal/mol) of *syn* isomers 6–9. In contrast, the *anti* compounds 10–13 formed only one of the ^+NH-OH , the $^+NH-O=$ or the ^+NH-OC interactions (Fig. 5b; Supplementary Fig. 35). Comparative NMR studies of the protonated form of one compound from each group (7 and 12) in DMSO- d_6 (polar high-dielectric medium surrogate for water) and CDCl₃ (lipophilic low-dielectric medium surrogate for the plasma-membrane interior) confirmed the findings from computational conformational analysis (Fig. 5c; Supplementary Figs. 36–38; Supplementary Tables 42–52). Thus, compound 7 exists in a rigid conformation in CDCl₃, with an intramolecular hydrogen bond network that shields the main polar functionalities and should facilitate cell permeability. However, compound 7 is present in multiple, more-polar conformations in DMSO- d_6 that should contribute to solubility, for example, in aqueous media (Supplementary Table 52). Compound 12, on the other hand, populates a mixture between two conformations, each exhibiting single $^+NH-O=$ or ^+NH-OC interactions in both CDCl₃ and DMSO- d_6 (2:1 in CDCl₃, approximately 1:1 in DMSO- d_6 , Supplementary Tables 42 and 48), explaining its lower permeability.

The two 1,5-triazoles 14 and 15, which are regioisomers of 7 and 12 with a smaller macrocyclic ring, had significantly higher $\log D_{7.4}$ values of 2.2 and one order of magnitude lower pK_a values ($pK_a = 7.5$ – 7.8 ; Supplementary Table 39). Consequently, they showed significantly higher cell permeabilities than 6–13. The surprising changes in properties could be explained by conformational analysis of 14 and 15. In these compounds, the protonated tertiary amine is less likely to generate an intramolecular hydrogen bonding network due to steric clashes within the macrocyclic ring (Fig. 5d). Since the charged form of 1,5-triazoles 14 and 15 is destabilized, the pK_a of the amine decreases, leading to an increase of $\log D_{7.4}$ and cell permeability.

Predicting cell permeability

Our models of macrocycle cell permeability, based on calculated molecular descriptors, predicted the average cell permeability for stereo- and regioisomeric clusters of DOS macrocycles with good accuracy (Figs. 2a and 6a). However, discrimination between stereo- and regioisomers is beyond the capabilities of high-throughput 2D-based QSPR models, as illustrated by the analysis of 6–13. This limitation can be partially overcome by including experimental data like $\log D_{7.4}$, but these data are often missing during the design process of new chemical entities. As conformational analysis proved successful in rationalizing the differences in passive permeability between the *syn* and *anti* sets of stereoisomers 6–13, we hypothesized that the approach could be broadly applicable to predict the rank order of permeability for other regio- or stereoisomeric clusters. It was therefore used to generate conformations for five additional clusters, followed by scoring of the overall polarity, the degree of intramolecular hydrogen bonding and general steric shielding of polar groups in the ensembles of low-energy conformations of each of

the isomers (Fig. 6b; Supplementary Table 53). Experimental data was kept blinded during scoring and analysis. Satisfactorily, this approach yielded rankings of passive permeability for the regio- and stereoisomers that were in excellent agreement with experimental data for four of the five clusters studied (Fig. 6c). For the outlier, cluster 3, both compounds were predicted to have equal and high cell permeability, whereas the experimental data revealed stereo-dependent permeability in the presence of efflux inhibitors.

DISCUSSION

The potential of macrocycles for modulating difficult targets in chemical biology and drug discovery^{2,4,8} is hampered by the limited understanding of which molecular properties allow macrocycles to be cell permeable and orally absorbed. To remedy this limitation, we generated an extensive data set with solubility, lipophilicity and cell permeability for >200 non-peptidic macrocycles at the border of and beyond the Ro5 using consistent experimental conditions. Cell permeability was determined in Caco-2 cell monolayers, which show excellent correlation with intestinal absorption, one of the key factors for oral bioavailability^{23–25}.

Over half of the >200 macrocycles had moderate to high cell permeability (Caco-2 P_{app} (A-B) $> 1 \times 10^{-6}$ cm/s), and 76% of these also had a high solubility ($> 50 \mu M$), which are properties in ranges that often allow oral administration. We show that high cell permeability is strongly and positively correlated with compound lipophilicity, whereas hydrogen bond acceptors and donors, charge and polarity limit permeability. These molecular properties also dominated permeability models of drugs in Ro5 space and in a comprehensive set of drugs and clinical candidates³ that reside even further into bRo5 space than the macrocycles investigated herein. We therefore conclude that analogous strategies for optimization of physicochemical properties may be used in all three chemical spaces, in contrast to previous suggestions²⁸. Certain polar functional groups were associated with poor passive permeability and/or significant efflux (1°- and 2°-amines, 1,4-triazoles, ureas, and sulfonamides; Fig. 4a), whereas more hydrophobic groups contributed to an increased permeability (phenyl, pyridine, 3°-amines, and isoxazoles). Some structural features, such as phenyl groups and tertiary amines, showed a synergistic improvement of cell permeability, while others, such as oxygen and nitrogen counts, showed antagonistic effects that reduced permeability even further than was expected from their combination. These structure–permeability relationships provide a confirmation for some colloquial knowledge, but also highlight additional motifs, and synergistic use of some motifs, for macrocycle design.

A significant fraction (>75%) of the macrocycles exhibited cellular efflux, which correlated to increased size, HBAs and polarity—in line with previous studies on non-macrocycles^{33–35}—as well as to known structural motifs for efflux transporters like sulfonamides^{30,31}. The size of macrocycles investigated herein is below the limit (radius of gyration $> 7 \text{ \AA}$) at which previous studies indicated passive permeability to be reduced³⁶; and we conclude that as the size of compounds increase, their permeability will be limited first by increased efflux, and then also by reduced passive permeability across the cell membrane. Efflux should therefore be monitored in development of macrocyclic drugs, but it should not be perceived as a major burden for oral absorption. No signs of active, transporter-mediated uptake across Caco-2 cells were observed, suggesting that active uptake from the intestine might not be of importance for *de novo*-designed macrocycles such as those investigated herein. This observation, and the need to monitor efflux, is in line with findings for drugs and clinical candidates in bRo5 space³.

Notably, our investigations revealed that stereo- and regiochemistry can have a significant influence on lipophilicity, as well as on passive permeability and cellular efflux, whereas their impact on solubility was lower. In >40% of the isomeric clusters studied, we found permeability differences above two-fold, suggesting that

both structural features and their 3D arrangements need to be adjusted carefully in the design of macrocyclic probes and drugs. Interestingly, *syn* isomer **7** displayed conformational flexibility (Fig. 5), leading to formation of dynamic intramolecular hydrogen bonds in an environment-dependent manner. This oscillation between hydrophilic, polarity exposed conformations and a 'membranophilic' conformation explains why **7** displays both high solubility and high cell permeability, while the more rigid and polar stereoisomer **12** has lower permeability. The dynamic relationship of conformation and environment and its impact on compound properties have been discussed previously^{3,4,14,24} and are generally assumed to explain the exceptional oral bioavailability of cyclosporin A³⁷. However, a combined environment-dependent effect on permeability and solubility, similar to that illustrated by macrocycles **7** and **12**, has only recently been exemplified through work on two carefully designed sets of cyclic peptides^{9,38}. We therefore conclude that appropriate conformational flexibility is a highly desirable property that contributes to both aqueous solubility and permeability in bRo5 space, and that macrocyclization is a preferred way to achieve this objective. Prospective incorporation of flexibility in macrocycle design remains to be demonstrated and may be challenging, as high flexibility has a pronounced negative effect on permeability and oral bioavailability³⁹.

We wanted to transform the insights obtained on determinants of macrocycle cell permeability into predictive tools that enable rational design in new regions of chemical space. Because of the lack of experimental data, available permeability models have poor coverage of macrocyclic and bRo5 chemical space⁴⁰. Satisfyingly, our quantitative structure–permeability models were externally predictive for randomly selected hold-out data (Fig. 2; **Supplementary Tables 33–38**), and predictions for stereo- and regioisomeric series were close to the series average (Fig. 6a), although individual isomers displayed up to ten-fold differences in experimentally determined permeabilities. Algorithms based on 2D molecular properties can thus be used to provide a first estimate of macrocycle permeability. However, we realized the need for taking the spatial properties of stereo- or regioisomers into consideration to achieve more accurate predictions. To this end we demonstrated the ability to rank series of stereo- or regioisomers through visual inspection of ensembles of calculated low-energy conformations (Fig. 6c), focusing on the extent of formation of intramolecular hydrogen bonds, shielding of polar functionalities and overall polarity. We highlight that, for conformationally flexible and stereochemically diverse compounds such as macrocycles, it is important to inspect not just a single minimized structure and its properties, since the ability to alternately shield or display polarity appears to be of general importance for cell permeability and aqueous solubility in beyond Ro5 space (Figs. 5 and 6). As throughput is currently limited, we envision detailed conformational analyses to be run on a focused set of proposed analogs, selected based on much faster and more general quantitative structure–permeability predictions using 2D property descriptors. Our approach, and those applied to cyclic peptides^{9,14}, should benefit chemical biology and drug discovery projects, especially for synthetically challenging macrocycles.

In conclusion, the comprehensive data set we have collected for >200 non-peptidic macrocycles, including sets of stereo- and regioisomers, offers a unique platform for investigating the molecular determinants of macrocycle physicochemical properties and cell permeability. It has allowed us to identify how functional groups and substituents impact permeability and to derive structure–permeability models that can be used to design cell-permeable macrocycles. Furthermore, we demonstrated that a fundamental understanding of the 3D conformations of macrocycles, their dynamics, and the corresponding impact on physicochemical properties like $\log D$, solubility, and pK_a improves permeability predictions even further. The procedure for conformational analysis described

herein, in combination with balancing of lipophilicity, HBD, HBA and MW through the use of structure–permeability models, now provides chemists with a rational approach to design cell-permeable and orally bioavailable non-peptidic macrocycles.

Received 24 October 2015; accepted 4 August 2016;
published online 17 October 2016

METHODS

Methods and any associated references are available in the [online version of the paper](#).

References

- Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
- 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.* **59**, 2312–2327 (2016).
- 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.* **21**, 1115–1142 (2014).
- Driggers, E.M., Hale, S.P., Lee, J. & Terrett, N.K. The exploration of macrocycles for drug discovery—an underexploited structural class. *Nat. Rev. Drug Discov.* **7**, 608–624 (2008).
- Surade, S. & Blundell, T.L. Structural biology and drug discovery of difficult targets: the limits of ligandability. *Chem. Biol.* **19**, 42–50 (2012).
- Mallinson, J. & Collins, I. Macrocycles in new drug discovery. *Future Med. Chem.* **4**, 1409–1438 (2012).
- Giordanetto, F. & Kihlberg, J. Macrocyclic drugs and clinical candidates: what can medicinal chemists learn from their properties? *J. Med. Chem.* **57**, 278–295 (2014).
- Villar, E.A. *et al.* How proteins bind macrocycles. *Nat. Chem. Biol.* **10**, 723–731 (2014).
- Bockus, A.T. *et al.* Probing the physicochemical boundaries of cell permeability and oral bioavailability in lipophilic macrocycles inspired by natural products. *J. Med. Chem.* **58**, 4581–4589 (2015).
- Hewitt, W.M. *et al.* Cell-permeable cyclic peptides from synthetic libraries inspired by natural products. *J. Am. Chem. Soc.* **137**, 715–721 (2015).
- Nielsen, D.S. *et al.* Improving on nature: making a cyclic heptapeptide orally bioavailable. *Angew. Chem. Int. Ed. Engl.* **53**, 12059–12063 (2014).
- Wang, C.K. *et al.* Rational design and synthesis of an orally bioavailable peptide guided by NMR amide temperature coefficients. *Proc. Natl. Acad. Sci. USA* **111**, 17504–17509 (2014).
- Rosenquist, Å. *et al.* Discovery and development of simeprevir (TMC435), a HCV NS3/4A protease inhibitor. *J. Med. Chem.* **57**, 1673–1693 (2014).
- Rezaei, T. *et al.* Conformational flexibility, internal hydrogen bonding, and passive membrane permeability: successful in silico prediction of the relative permeabilities of cyclic peptides. *J. Am. Chem. Soc.* **128**, 14073–14080 (2006).
- Thansandote, P. *et al.* Improving the passive permeability of macrocyclic peptides: balancing permeability with other physicochemical properties. *Bioorg. Med. Chem.* **23**, 322–327 (2015).
- Wang, C.K. *et al.* Exploring experimental and computational markers of cyclic peptides: charting islands of permeability. *Eur. J. Med. Chem.* **97**, 202–213 (2015).
- Schreiber, S.L. Organic chemistry: molecular diversity by design. *Nature* **457**, 153–154 (2009).
- Lovering, F. Escape from flatland 2: complexity and promiscuity. *Med. Chem. Commun.* **4**, 515–519 (2013).
- Fitzgerald, M.E. *et al.* Build/couple/pair strategy for the synthesis of stereochemically diverse macrolactams via head-to-tail cyclization. *ACS Comb. Sci.* **14**, 89–96 (2012).
- Marcaurelle, L.A. *et al.* An aldol-based build/couple/pair strategy for the synthesis of medium- and large-sized rings: discovery of macrocyclic histone deacetylase inhibitors. *J. Am. Chem. Soc.* **132**, 16962–16976 (2010).
- Comer, E. *et al.* Fragment-based domain shuffling approach for the synthesis of pyran-based macrocycles. *Proc. Natl. Acad. Sci. USA* **108**, 6751–6756 (2011).
- Waring, M.J. Lipophilicity in drug discovery. *Expert Opin. Drug Discov.* **5**, 235–248 (2010).
- Artursson, P. & Karlsson, J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem. Biophys. Res. Commun.* **175**, 880–885 (1991).
- Artursson, P., Palm, K. & Luthman, K. Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv. Drug Deliv. Rev.* **22**, 67–84 (1996).

25. Hubatsch, I., Ragnarsson, E.G. & Artursson, P. Determination of drug permeability and prediction of drug absorption in Caco-2 monolayers. *Nat. Protoc.* **2**, 2111–2119 (2007).
26. Giacomini, K.M. *et al.* Membrane transporters in drug development. *Nat. Rev. Drug Discov.* **9**, 215–236 (2010).
27. Matsson, P., Pedersen, J.M., Norinder, U., Bergström, C.A. & Artursson, P. Identification of novel specific and general inhibitors of the three major human ATP-binding cassette transporters P-gp, BCRP and MRP2 among registered drugs. *Pharm. Res.* **26**, 1816–1831 (2009).
28. Heinis, C. Drug discovery: tools and rules for macrocycles. *Nat. Chem. Biol.* **10**, 696–698 (2014).
29. Matsson, P. *et al.* Exploring the role of different drug transport routes in permeability screening. *J. Med. Chem.* **48**, 604–613 (2005).
30. Raub, T.J. P-glycoprotein recognition of substrates and circumvention through rational drug design. *Mol. Pharm.* **3**, 3–25 (2006).
31. Zamek-Gliszczyński, M.J. *et al.* The important role of Bcrp (Abcg2) in the biliary excretion of sulfate and glucuronide metabolites of acetaminophen, 4-methylumbelliferone, and harmol in mice. *Mol. Pharmacol.* **70**, 2127–2133 (2006).
32. Ferreira, R.J., Ferreira, M.J.U. & dos Santos, D.J.V.A. Molecular docking characterizes substrate-binding sites and efflux modulation mechanisms within P-glycoprotein. *J. Chem. Inf. Model.* **53**, 1747–1760 (2013).
33. Desai, P.V., Raub, T.J. & Blanco, M.J. How hydrogen bonds impact P-glycoprotein transport and permeability. *Bioorg. Med. Chem. Lett.* **22**, 6540–6548 (2012).
34. Hitchcock, S.A. Structural modifications that alter the P-glycoprotein efflux properties of compounds. *J. Med. Chem.* **55**, 4877–4895 (2012).
35. Kuhn, B., Mohr, P. & Stahl, M. Intramolecular hydrogen bonding in medicinal chemistry. *J. Med. Chem.* **53**, 2601–2611 (2010).
36. 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.* **52**, 882–890 (2012).
37. 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.* **2**, 669–674 (2011).
38. Bockus, A.T. *et al.* 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.* **58**, 7409–7418 (2015).
39. Varma, M.V.S. *et al.* Physicochemical space for optimum oral bioavailability: contribution of human intestinal absorption and first-pass elimination. *J. Med. Chem.* **53**, 1098–1108 (2010).
40. Mathiowetz, A.M., Leung, S.S. & Jacobson, M.P. in *Macrocycles in Drug Discovery* Vol. 40 (ed. Levin, J.) 367–397 (Royal Soc. Chem. Press, Cambridge, 2014).
41. Bickerton, G.R., Paolini, G.V., Besnard, J., Muresan, S. & Hopkins, A.L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **4**, 90–98 (2012).

Acknowledgments

This work was funded by a postdoctoral fellowship at AstraZeneca R&D Gothenburg (B.O.), the Carl Trygger Foundation (P. Matsson), the Swedish Research Council (grant no. 2822; P.A.) and the NIGMS-sponsored Center of Excellence in Chemical Methodology and Library Development (Broad Institute CMLD; P50 GM069721). ChemAxon and Simulations Plus are graciously acknowledged for providing access to the Instant JChem and ADMET Predictor software, respectively. We thank J. Wernevik, O. Hedge and R. Svensson for determining logD and pK_a data, respectively, and J. Holmgren, L. Fredlund and C. Vedin for assistance with Caco-2 cell measurements. We also thank C. Mulrooney and J. Ulander for helpful discussions pertaining to computational experiments.

Author contributions

B.O. and P. Matsson made major contributions to this work. J.K., C.H., M.A.F. and P.A. initiated the project. J.K., J.R.D., C.H., M.W.D.P., P.A., P. Matsson and B.O. designed experiments. J.K., J.R.D., C.H., M.W.D.P. and P.A. supervised the work. J.R.D., M.A.F., M.D.L. and G.M. were part of a team at the Broad Institute which designed and synthesized the macrocycles. B.O. performed biochemical experiments and cell assays. S.E.J. measured solubility. B.O. and R.J.L. performed NMR studies. B.C.D. assembled the beyond rule of 5 data set. C.T., P. McCarren, P. Matsson and B.O. performed conformational analysis. P. Matsson built the chemical networks for compound selection and PLS and RF multivariate regression models. U.N. built SVM and RF regression models. C.T. predicted cell permeability. P. Matsson, B.O. and B.C.D. analyzed the data. J.K., B.O., P. Matsson and J.R.D. prepared the manuscript, with feedback and contributions from the other authors.

Competing financial interests

The authors declare competing financial interests: details accompany the [online version of the paper](#).

Additional information

Any supplementary information, chemical compound information and source data are available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to J.K. or J.R.D.

ONLINE METHODS

Compound selection. All macrocycles profiled in this study have been reported in PubChem, and details of their synthesis have been reported elsewhere^{19–21}. Although it was neither a selection criterion nor an objective of the present studies, we found the selected compounds to have yielded active hits against *Trypanosoma cruzi*, *Plasmodium falciparum*, class I histone deacetylase (HDAC) and the Sonic Hedgehog signaling pathway, to name a few. A total of 77 activities in 290 assays have so far been reported in PubChem for the selected macrocycles, which illustrates their ability to modulate biological systems.

A representative compound subset covering chemical property space and providing stereochemical diversity was selected using a chemical similarity network approach. Compounds were connected in a network based on their pairwise similarity in molecular weight, molecular volume, calculated octanol–water partition coefficient, the surface areas of polar and non-polar atoms, and the numbers of hydrogen bond donors and acceptors. Network preparation was performed using the igraph R package version 0.7.1, with visualization in R 3.1.1 (<http://www.r-project.org>) using custom code. Representative diverse compounds from each structural series were selected by picking the medoids from compound clusters across the entire network along with stereo- or regioisomers of these. The full complement of stereo- and regioisomers were included for a structurally diverse subset of the compound clusters; for all other selected cluster medoids, a single stereo- or regiocenter was varied at a time such that, collectively, the selected compounds included all possible isomeric variants. In order to cover the maximal range of molecular shapes for each sublibrary (sphere-, rod- and disc-like structures)⁴², four compounds covering all combinations of high/low normalized principal moment of inertia ratios were also included (i.e., the shape extremes in each sublibrary).

Solubility measurements. Solubility was determined in PBS pH 7.4 containing 1% DMSO. Compounds were prepared at 100 μ M by diluting 10 mM DMSO stock solutions with PBS in triplicate. Compounds were allowed to equilibrate at room temperature for 18 h on a 750 r.p.m. vortex shaker. Stir Stix (V&P Scientific) were included in each preparation to limit compound aggregation⁴³. After equilibration, samples were centrifuged to remove precipitate and the supernatants were analyzed by UPLC-MS (Waters) with SIR detection on a single quadrupole mass spectrometer. The response was fitted to a two-point standard curve prepared in DMSO.

logD measurements. logD was measured with a scaled-down version of the traditional shake flask technique, using UPLC-MS to measure the relative octanol and aqueous concentrations. 1-Octanol (HPLC grade $\geq 99\%$; Sigma-Aldrich) and a 10 mM phosphate buffer [$\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$ (p.a. grade; Merck) and $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ (p.a. grade; Merck)] were used. Equal parts of buffer and 1-octanol were vigorously mixed in a separation funnel three times (at least 15 min between each mixing) to saturate the solutions. Phases were separated overnight before experiments. Compounds were assayed in pools of eight, and four dilutions of buffer and octanol samples were analyzed and evaluated for logD calculation. All liquid transfers were performed with a Beckman Biomek FX robot and samples were analyzed on a fast-scanning triple-quadrupole mass spectrometer (Waters Micromass TQD with MassLynx 4.1) coupled to a Waters Acquity Ultra Performance LC using an Acquity UPLC HSS T3 1.8 μ m, 2.1 \times 50 mm or an Acquity UPLC BEH C18 1.7 μ m, 2.1 \times 50 mm column.

pK_a measurements. Acid dissociation constants (pK_a) were determined UV-metrically with a Sirius T3 instrument equipped with a D-PAS (Dip Probe Absorption Spectroscopy) lamp from Sirius Analytical Ltd. DMSO compound solutions (10 mM, 2–5 μ L depending on the compound) were added to 25 μ L phosphate buffer to ensure adequate buffering capacity during the titration, and ionic-strength-adjusted water was added (1.5 mL). A pH-metric titration from low to high pH was performed, establishing a titration curve based on UV-Visible spectral detection with the D-PAS technique. The electrode was calibrated using a blank titration from pH 1.8 to 12.0 before every individual determination. Measurements were performed under argon to minimize effects from dissolved CO₂. Precipitation was continuously monitored at 500 nm; this was apparent for the more insoluble compounds. The temperature was controlled throughout the experiment at 25 \pm 1 $^\circ$ C.

Caco-2 cell monolayer permeability measurements. Transepithelial transport across Caco-2 cell monolayers was determined in an automated fashion using a Tecan Freedom EVO 200 equipped with a TeMO 96 pipetting head and a transepithelial electrical resistance (TEER) measurement station. Caco-2 cells (obtained from ATCC; confirmed mycoplasma-free) were cultured in DMEM supplemented with 10% FCS (Invitrogen) on Costar 24-well cell culture cluster plates (polycarbonate membrane, 0.4 μ m pore size; Corning), for 15–19 d before experiments. Prior to the transport assay, cells were washed with pre-warmed HBSS supplemented with 25 mM HEPES (HBSS–HEPES pH 7.4) to remove the culture medium. After 15 min equilibration, TEER was measured to determine acceptance of the cell plates into the assay. TEER and Lucifer-yellow leakage were also determined after each transport experiment to verify monolayer integrity throughout the study.

Permeability was determined in both the A-to-B and B-to-A directions, with HBSS–HEPES pH 7.4 on both sides of the cell monolayer. Compounds were assayed in duplicate at 10 μ M concentration in a shaking incubator at 480 r.p.m. at 37 $^\circ$ C. All compounds were assayed in semi-high-throughput mode with sampling from the receiver side at 60 min and from the donor side at 0 and 60 min in both transport directions. In addition, selected compounds were assayed in more detail, with sampling from the receiver side at 45 and 120 min for experiments in the A-to-B direction and at 20 and 60 min for the B-to-A direction, using the same buffer conditions as above. For these compounds, permeability was also measured in the presence of 50 μ M quinidine, 20 μ M sulfasalazine and 30 μ M benzbramarone to inhibit major ABC efflux transporters. Transporter-specific inhibitors were also used for compounds 1–16 (Supplementary Table 40). All inhibitors were purchased from Sigma-Aldrich and were of analytical grade. Mass recovery was determined in all experiments and was generally good (average 95%).

Apparent permeability coefficients were calculated from the rates of compound appearance in the receiver according to

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

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 were calculated as the ratio of permeability coefficients in the B-to-A and A-to-B directions. The non-effluxed control metoprolol has an efflux ratio of 0.5 in this assay setup (Supplementary Table 40). The specific contribution from efflux transporters to the apparent permeability, correcting for the impact of passive background permeability, was calculated from efflux ratios according to the equation⁴⁴:

$$P_{\text{app}}(\text{Efflux}) = 2 \times P_{\text{app}}(\text{Passive}) \times (P_{\text{app}}(\text{B-A})/P_{\text{app}}(\text{A-B}) - 1).$$

NMR measurements. A Bruker Avance III 500 MHz system equipped with a 5 mm cryogenically cooled QNP probe was used to analyze the compounds (protonated with TFA if relevant) in DMSO- d_6 (Sigma-Aldrich; 99.9% deuteration) or CDCl_3 (Sigma-Aldrich, 99.8% deuteration) at 298 K. Spectral assignment was typically performed using 0.5 mg substance, dissolved in 140 μ L DMSO- d_6 and transferred to a 3 mm tube. ^1H and ^{13}C resonances were assigned using standard 1D and 2D Bruker pulse sequences (TopSpin version 3.0). H-H-COSY ⁴⁵, C-H-HSQC ⁴⁶ and C-H-HMBC ⁴⁷ experiments were performed to determine ^1H and ^{13}C connectivities. H-H-ROESY ⁴⁷ (200 ms spinlock time) was used to obtain qualitative intramolecular distance information. ^1H and ^{13}C chemical shifts were referenced relative to residual solvent signals at 2.50 and 39.52 p.p.m. (DMSO) and 7.26 and 77.16 p.p.m. (CDCl_3). In the protonated form, 7 and 12 (Fig. 5) exist as two diastereomers at the quaternary ammonium group. The observed ROESY interactions and solvent-dependent differences in chemical shift were classified as strong or weak, and key long-range interactions were compared with those observed in the 60 (30 S and 30 R diastereomers) lowest-energy structures (calculated with LowModeMD and a distance-dependent electrostatic model) to identify the conformations present in the respective solvent (Supplementary Tables 42–52; Supplementary Figs. 35–38).

Hydrogen bond strength was quantitatively assessed by calculating the hydrogen bond acidity A^{H} based on the chemical shift differences of hydroxyls and/or amines in DMSO versus CDCl_3 (Supplementary Table 42).

$$A = 0.0065 + 0.133 \times \Delta\delta \text{ with } \Delta\delta = \delta(\text{DMSO}) - \delta(\text{CDCl}_3)$$

Hydrogen bond donor and acceptor counts. To allow direct comparisons with the Ro5, hydrogen bond donors (OH + NH) and acceptors (N + O) were calculated as in the original Ro5 publication¹. Software-dependent differences in the assignment of hydrogen bond functionality are shown in **Supplementary Figures 7–9**.

Quantitative structure–property relationship (QSPR) modeling. Partial least-squares (PLS), support vector machine (SVM), and random forest (RF) regression were used to develop computational models of cell permeability and transporter-mediated efflux, using molecular descriptors from Dragon 6 (Taleté) and ADMET Predictor, version 6.0 (SimulationsPlus), or substructural features identified with Instant JChem version 6.0.2.943 (ChemAxon) using custom SMARTS strings.

For PLS, model parameters and variable selection were optimized using an automated double-loop cross validation procedure to provide an unbiased estimate of prediction accuracy^{49,50}; model performance was iteratively evaluated on data withheld from the model training. Variable selection was not performed for substructure-based models. The entire double-loop procedure was repeated 100 times for different random partitionings of the data set to calculate confidence intervals of prediction accuracy estimates and model parameters. Statistical significance of the final models was assessed using permutation of the dependent variable; permutation $P < 0.0001$ for all models. PLS models were developed using either molecular properties or substructural features in the descriptor set. In addition, PLS models were also trained using substructure counts as descriptors and including all pairwise cross terms among the descriptors. In such models, significant cross terms ($A \times B$) indicate that the impact of descriptor A depends on the value of descriptor B (and vice versa).

SVM models were derived for 100 random training-test set splits, using LIBSVM version 3.11 with epsilon-SVR and the RBF-kernel⁵¹. Default parameters were used, apart from γ and c , which were set to 0.002 and 50, respectively. RF models were derived using the random forest R package v4.6-12 (ref. 52), with default settings except for the number of trees in the forests (100 trees).

For visualization of model similarity, intercorrelated molecular descriptors were clustered using hierarchical clustering with complete linkage, and the maximal absolute model coefficient in each cluster was used to define its importance. Inter-model similarity was calculated as the Pearson correlation coefficient between model coefficients for molecular descriptors included in either of the compared models.

Conformational analysis. Macrocyclic conformations were calculated using LowModeMD sampling in MOE 2013.0802. The energy window for keeping structures was set to 15 kcal/mol, the iteration limit to 10,000, and the MM iteration limit to 2,000. Minimization was performed using the MMFF94x force field with a distance dependent dielectric model as an implicit solvation term. The lowest energy conformation from selected structural interaction clusters were optimized using density functional theory B3LYP-d3, 6-31**++ basis set in Jaguar (Maestro 9.8). For the calculation of Boltzmann weighted microstate probabilities (**Fig. 6b**), eight microstates were defined based on the patterns of

intramolecular hydrogen bonds among the protonated amine, carbonyl and hydroxyl functions: (1) NH–OH–O=C, (2) NH–OH–OC, (3) NH–OH, (4) NH–O=C, (5) OH–O=C, (6) NH–OC, (7) OH–OC, (8) no detectable hydrogen bonds. A maximum distance of 2.6 Å between donor and acceptor was used to define the presence of hydrogen bonds. The probabilities of microstates in the statistical ensemble were calculated with

$$Z = \sum_i e^{-\beta \cdot E_i}$$

$$P_i = \frac{1}{Z} \cdot e^{-\beta \cdot E_i}$$

$$\beta = \frac{1}{k_B T}$$

where Z is the partition function, P_i the probability of state i , k_B is the Boltzmann constant, E is the average energy of the system in the respective microstate and T is the temperature.

3D structure figures were prepared with the PyMol Molecular Graphics System (Version 1.7.4, Schrödinger), 2D schemes using ChemDraw Ultra (Version 11.0.2, CambridgeSoft).

Statistical analysis. Statistical analyses were performed in R 3.1.1 (<http://www.r-project.org>). Student's t -test was used for statistical comparisons, with Bonferroni-corrected P values of <0.05 considered significant.

42. Sauer, W.H.B. & Schwarz, M.K. Molecular shape diversity of combinatorial libraries: a prerequisite for broad bioactivity. *J. Chem. Inf. Comput. Sci.* **43**, 987–1003 (2003).
43. Alelyunas, Y.W., Liu, R., Pelosi-Kilby, L. & Shen, C. Application of a Dried-DMSO rapid throughput 24-h equilibrium solubility in advancing discovery candidates. *Eur. J. Pharm. Sci.* **37**, 172–182 (2009).
44. Kalvass, J.C. & Pollack, G.M. Kinetic considerations for the quantitative assessment of efflux activity and inhibition: implications for understanding and predicting the effects of efflux inhibition. *Pharm. Res.* **24**, 265–276 (2007).
45. von Kienlin, M., Moonen, C.T.W., van der Toorn, A. & van Zijl, P.C.M. Rapid recording of solvent-suppressed 2D COSY spectra with inherent quadrature detection using pulsed field gradients. *J. Magn. Reson.* **93**, 423–429 (1991).
46. Willker, W., Leibfritz, D., Kerssebaum, R. & Bermel, W. Gradient selection in inverse heteronuclear correlation spectroscopy. *Magn. Reson. Chem.* **31**, 287–292 (1993).
47. Marion, D. Rotating frame nuclear overhauser effect: a practical tool for the ¹H NMR study of peptides in solution. *FEBS Lett.* **192**, 99–103 (1985).
48. Abraham, M.H. *et al.* An NMR method for the quantitative assessment of intramolecular hydrogen bonding: application to physicochemical, environmental, and biochemical properties. *J. Org. Chem.* **79**, 11075–11083 (2014).
49. Freyhult, E. *et al.* Unbiased descriptor and parameter selection confirms the potential of proteochemometric modelling. *BMC Bioinformatics* **6**, 50 (2005).
50. Mateus, A., Matsson, P. & Artursson, P. Rapid measurement of intracellular unbound drug concentrations. *Mol. Pharm.* **10**, 2467–2478 (2013).
51. Chang, C.C. & Lin, C.J. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 27:1–27:27 (2011).
52. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).