KEYNOTE (GREEN)

Exploring the chemical space of orally bioavailable PROTACs

Giulia Apprato [1],
Vasanthanathan Poongavanam [2],
Diego Garcia Jimenez [1], Yoseph Atilaw [2],
Mate Erdelyi [2], Giuseppe Ermondi [1],
Giulia Caron [1,*], Jan Kihlberg [2,*]

[1] Department of Molecular Biotechnology and Health Sciences,
University of Torino, Piazza Nizza 44bis, 10126 Torino, Italy
[2] Department of Chemistry – BMC, Uppsala University, SE-75123 Uppsala, Sweden

A principal challenge in the discovery of proteolysis targeting chimeras (PROTACs) as oral medications is their bioavailability. To facilitate drug design, it is therefore essential to identify the chemical space where orally bioavailable PROTACs are more likely to be situated. To this aim, we extracted structure-bioavailability insights from published data using traditional 2D descriptors, thereby shedding light on their potential and limitations as drug design tools. Subsequently, we describe cutting-edge experimental, computational and hybrid design strategies based on 3D descriptors, which show promise for enhancing the probability of discovering PROTACs with high oral bioavailability.

Giulia Apprato is a PhD student at the University of Turin, Italy, in the CASSMedChem research team. She completed a master's degree in molecular biotechnology - diagnostics and drug discovery at the university, and she is currently focusing on setting up drug discovery strategies to improve the current preclinical 'beyond rule of five' (bRo5) pipelines. Her most recent result is a new *in silico* tool for modeling proteolysis targeting chimera (PROTAC) ternary complexes.

Vasanthanathan Poongavanam is a senior researcher in the Department of Chemistry -BMC, Uppsala University, Sweden. Before starting at Uppsala University in a post-doctoral position with Jan Kihlberg in 2016, he was a postdoctoral fellow at the University of Vienna, Austria, and at the University of Southern Denmark in Odense. He obtained his Ph.D. in medicinal chemistry at the University of Copenhagen, Denmark. He has published more than 60 scientific articles, including reviews and book chapters. His research focuses on applying computational chemistry and artificial intelligence methods to drive drug discovery projects, in particular for understanding the molecular properties of molecules beyond the Ro5 space, including macrocycles and PROTACs.

Diego Garcia Jimenez is currently a researcher in pharmaceutical and biomolecular sciences at the University of Turin, Italy. His primary research focus lies in investigating the molecular properties of drug candidates within the bRo5 chemical space, with a specific emphasis on PROTACs and macrocycles. Recently, his interest was extended to the study of molecular chameleonicity as a tool to improve the absorption, distribution, metabolism and excretion (ADME) profile of bRo5 compounds.

Yoseph Atilaw is a senior research scientist at AstraZeneca R&D in Gothenburg, Sweden. His main research interest is in developing new methods to determine solution conformation using nuclear magnetic resonance (NMR) and vibrational circular dichroism (VCD) spectroscopy. He did his postdoctoral research at Uppsala University, Sweden, with Jan Kihlberg and Mate Erdelyi on the determination of solution conformations for drug-like molecules. He did his PhD at the University of Nairobi, Kenya, under the supervision of Abiy Yenesew in the field of natural products chemistry. His work has been published in a number of peer-reviewed journals.

* Corresponding authors. Caron, G. (giulia.caron@unito.it), Kihlberg, J. (jan.kihlberg@kemi.uu.se).

Mate Erdelyi has been a professor in organic chemistry at Uppsala University, Sweden, since 2017. His passion is NMR spectroscopy. He enjoys investigating conformational ensembles, halogen bonding, bioactive herbal natural products and developing methods to understand protein-ligand interactions, as well as weak non-covalent forces. He graduated from Semmelweis University in Budapest, Hungary, in 1999, obtained his PhD at Uppsala University, was trained as postdoctoral researcher at the University of California San Diego and at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany. He initiated his independent research at the University of Gothenburg, Sweden, in 2008.

Giuseppe Ermondi is an associate professor of medicinal chemistry in the Department of Molecular Biotechnology and Health Sciences at the University of Turin, Italy. The main focus of his scientific research is the application of computational methods for the prediction of bRo5 structural features responsible for the drug metabolism and pharmacokinetics, and target modulation. Recently, he has become interested in rare diseases and the computational methods for the identification of pathogenic missense mutations at the structural level. He has authored more than 110 papers and chapters in books.

Giulia Caron is an associate medicinal chemistry professor in the Molecular Biotechnology and Health Sciences Department at the University of Turin, Italy. She is mainly interested in property-based drug discovery applied to the bRo5 chemical space. She has a specific focus on the computational prediction and experimental determination of ionization, lipophilicity, polarity and chameleonicity. Moreover, she is involved in the integration of intramolecular-hydrogen-bond considerations in PROTAC design. She has an extensive publication record, with more than 120 papers and book chapters to her name.

Jan Kihlberg has been a chair in organic chemistry at Uppsala University, Sweden, since 2013. His key research interests involve understanding what properties convey cell permeability, aqueous solubility and target binding to drugs in the bRo5 space, and translating this knowledge into guidelines for design. He has published more than 200 peer-reviewed publications and book chapters. Before moving to Uppsala, he spent ten years at AstraZeneca R&D in Gothenburg, Sweden, seven of which were as director of medicinal chemistry.

Introduction

Induction or strengthening of complex formation between two proteins has the potential to modulate a vast number of biological processes, thereby providing a major increase of druggable target space.[p1] Complex inducers or stabilizers include molecular glues that inhibit the function of one of the proteins in the complex, as well as different heterobifunctional compounds that mediate modulation of the post-translational modifications of target proteins or induce their degradation by the proteasome or in lysosomes. Proteolysis targeting chimeras (PROTACs) are heterobifunctional compounds consisting of a ligand for a target protein connected via a linker to another ligand that binds to an E3 ubiquitin ligase.[p2] PROTAC-induced ternary complex formation results in ubiquitination and subsequent degradation of the target protein by the proteasome. The majority of PROTACs are based either on a cereblon (CRBN) or a Von Hippel-Lindau (VHL) E3 ligase ligand.[p3]

Although the guidelines of Lipinski's rule of 5 (Ro5)[p4] and Veber's rule[p5] were deduced by analysis of 'traditional' small molecule drugs and drug candidates, these guidelines also provide a first indication of the likelihood that other chemical modalities display oral bioavailability. The heterobifunctional structure of PROTACs locates them in the beyond rule of 5 (bRo5) chemical space: that is, in a chemical space beyond that defined by the Ro5 and Veber's rule.[p6],[p7] Specifically, most PROTACs are found close to or beyond the outer borders of oral druggable space derived from analysis of drugs, clinical candidates and compounds in lead optimization,[p8],[p9] with CRBN-based PROTACs being closer to traditional drugs than VHL-based ones.[p10],[p6],[p7] Consequently, it is challenging to discover orally bioavailable PROTACs, and parenteral administration might be the preferred route of administration for some of them.[p11]

Convenience, high patient adherence and cost-effectiveness make oral administration the preferred method of drug delivery.[p12] The oral bioavailability (F%) quantifies the fraction of the orally administered dose that reaches the systemic circulation after passing the liver, whereas the oral absorption defines the fraction that reaches the liver. Solubility, cell permeability and first-pass metabolism in the liver are the three most important determinants of oral bioavailability. A high oral bioavailability reduces the dose required to achieve the desired pharmacological effect and reduces patient-to-patient variability, both of which decrease the likelihood of side effects and toxicity.[p13] In clinical trials, low oral bioavailability frequently leads to the failure of drug candidates to reach the market. Nevertheless, the experimental determination of bioavailability is costly and time-consuming, allowing only a few compounds to be investigated. Consequently, it is crucial to (i) extract insights from existing bioavailability data and (ii) establish guidelines from *in silico* and *in vitro* assays that increase the likelihood of discovering compounds with a high bioavailability. This is of particular significance for PROTACs, where oral bioavailability is anticipated to be the primary obstacle to their development as oral medications.

The research leading to the discovery of orally bioavailable PROTACs, as well as medicinal chemistry approaches for the optimization of the three components of PROTACs, has been summarized in recent publications.[p14],[p15],[p16] Herein we discuss how the chemical space of orally bioavailable PROTACs can be defined in a more consistent way. We also propose that the use of conformation-dependent 3D descriptors, chromatographically determined physicochemical descriptors and *in silico* predictions should contribute to the more effective discovery and optimization of oral PROTACs. To meet these aims, we have generated a data set of PROTACs that display oral bioavailability in rodents and review recent structure–property reports for selected PROTACs. We hope that the insight provided herein will improve the design of oral PROTACs in the future.

PROTACs in clinical trials

In the first half of 2023, 20–25 PROTACs were reported to be in clinical trials or approved for entering clinical trials.[p1],[p11] The chemical structures have been disclosed for eight of them: seven are based on different types of CRBN E3-ligase ligands, and the eighth is the VHL PROTAC DT-2216 (Figure S1 in the supplementary material). All seven CRBN PROTACs are administered orally in the clinical studies and the oral bioavailability in mice has been reported for four of them (all have F >30%). By contrast, DT-2216 has a very low bioavailability (F <0.03%, mouse) and is administered by intravenous infusion. This difference in bioavailability agrees well with expectations from the descriptors of Lipinski's Ro5[p4] and Veber's rule.[p5] The VHL PROTAC DT-2216 is larger [higher molecular weight (MW)], more lipophilic [higher calculated partition coefficient between water and octanol (cLogP)], more polar [higher number of hydrogen bond donors (HBDs) and hydrogen bond acceptors (HBAs), and larger topological polar surface area (TPSA)] and more flexible [higher number of rotatable bonds (NRotBs)] than the four CRBN PROTACs. However, a data set of five PROTACs is too small to provide deeper insight into what properties allow oral bioavailability.

PROTACs often have high potency and have been found to have a catalytic mode of action,[p17] properties that might allow oral administration at low oral bioavailabilities and/or the use of low doses. Interestingly, these hypotheses are not supported by the once-daily oral doses of the three CRBN PROTACs that have advanced into phase II or III studies. They are all administered in high doses: ARV-766 (phase II) at 100 and 300 mg,[p18] ARV-110 (phase II) at a recommended dose of 420 mg[p19] and ARV-471 (phase III) at 200 mg,[p20] while the bioavailabilities are >30% for ARV-110 and ARV-471 (data have not been reported for ARV-766). However, data for additional PROTACs are required before more general conclusions can be drawn.

Overview of orally bioavailable PROTACs

To provide a larger data set of orally bioavailable PROTACs with disclosed chemical structures, we searched PubMed in August 2023 and then manually analyzed the retrieved articles. We found that oral bioavailability had been determined for 55 PROTACs in either mice or rats (Figure 1). These data originate from different laboratories that might have used different peroral formulations and are also subject to variation between the two species. However, keeping these uncertainties in mind, the 55 PROTACs provide a useful overview of the current oral PROTAC space and of how the structure of PROTACs influences their oral bioavailability.

Most of the retrieved PROTACs (47 out of 55) are based on a CRBN E3-ligase ligand, but seven rely on a VHL ligand and one is based on X-linked inhibitor of apoptosis protein (XIAP) (Figure 1a,b). A large proportion of the 55 PROTACs (N = 20) display a very low oral bioavailability in rodents (F ≤5%) and are not likely to allow oral administration to patients (Figure 1a). This bin comprises several CRBN PROTACs, all but two of the VHLs and the single XIAP PROTAC. However, other CRBN PROTACs show bioavailabilities ranging up to 89%, and a large number fall in the 20–40% bin, including one of the VHL PROTACs (ACBI2).[p21] The dominance of CRBN over VHL PROTACs among PROTACs that have satisfactory bioavailability has been

reported previously by AstraZeneca in an analysis of oncology projects focused on undisclosed targets.[p22] The overwhelming majority of our set of PROTACs are also directed towards oncology targets, with more than half being aimed at the androgen and estrogen receptors (Figure 1b). Different kinases constitute another large group of targets for the current set of PROTACs.

2D descriptors of chemical space and their limitations

A recent analysis of the oral absorption of 1,806 PROTACs, for which structures were not disclosed, suggested upper limits for the six 2D descriptors of Lipinski's[p4] and Veber's rules for PROTACs (e.g., MW ≤950 Da).[p16] Calculation of the six descriptors for commonly used CRBN and VHL ligands, as well as representative linkers, provided an estimate of the 'budget' remaining for the target binding ligand. This exercise highlighted that discovery of orally bioavailable VHL PROTACs constitutes a major challenge and that oral CRBN PROTACs are not trivial. The discovery of the PROTAC ACBI2 (F: 22%), however, provides an example of how controlling 2D physicochemical descriptors and iterative structure-based design can be used in the discovery of orally bioavailable VHL PROTACs.[p21] In ACBI2, the number of hydrogen bond donors in the ligand for the target, SMARCA2, was reduced to a minimum, while the linker was kept short and its composition tuned to combine optimal SMARCA2 degradation and pharmacokinetic properties. Determination of several crystal structures of SMARCA2-ligand complexes and ternary SMARCA2-PROTAC-VHL complexes were instrumental for optimization of the SMARCA2 ligand, the exit vector for the linker on the VHL ligand, and the length and structure of the linker.

Ideally, one would like to identify the chemical space in which orally bioavailable PROTACs are more likely to be found to enable drug design efforts. Therefore, we performed a principal component analysis (PCA) to investigate the chemical space populated by the 55 orally bioavailable PROTACs, and to compare this space to that of the orally bioavailable drugs in DrugBank[p23] and to all PROTACs in the PROTAC-DB 2.0 database[p24] (Figure 1c,d). The PCA was based on the six descriptors of Lipinski's Ro5 and Veber's rule, the number of carbon atoms (nC) and Kier's flexibility index (Φ).[p25],[p26] The chemical space of the three different sets of compounds is described by descriptors of: (i) size (MW and nC) and flexibility (NRotB and Φ, all of which are highly correlated; (ii) the three descriptors of polarity (HBD, HBA and TPSA), which are also highly correlated; and (iii) the calculated lipophilicity (cLogP). Unsurprisingly, the PROTACs in the PROTAC-DB were in general larger and more flexible, and contained more polar groups than the orally bioavailable drugs in DrugBank.[p6],[p7],[p10] CRBN PROTACs from the PROTAC-DB were found in the physicochemical descriptor space adjacent to or slightly overlapping with the oral drugs, whereas PROTACs with VHL ligands and E3-ligase ligands of other types resided in the physicochemical space further away from the orals. The orally bioavailable CRBN and VHL PROTACs in our data set are found towards the more drug-like end of the physicochemical space of the CRBN and VHL PROTACs of the PROTAC-DB (Figure 1d). Importantly, low to high bioavailabilities can still be obtained for PROTACs far outside the drug-like space defined by Lipinski's Ro5 and Veber's rule. CRBN PROTACs that display bioavailabilities from 5% to 89%

KEYNOTE (GREEN)
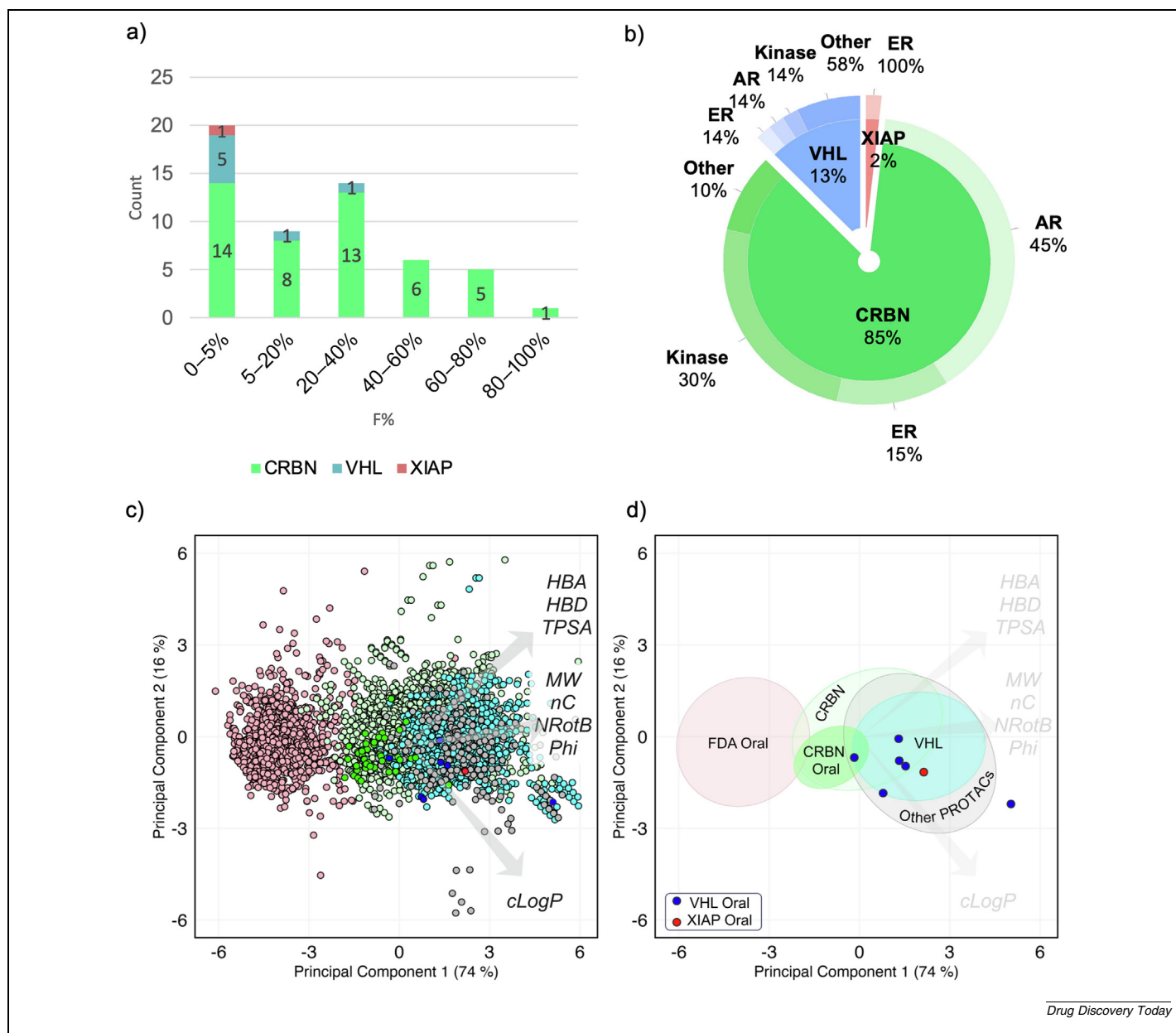
FIGURE 1

(a) Binned oral bioavailabilities (F%) for the members of the set of 55 PROTACs for which the oral bioavailability has been determined in mice or rats (data retrieved on August 31, 2023). (b) Targets for the orally bioavailable PROTACs displayed by their E3 ligase ligand. (c) Score plot of the first two principal components from a PCA, which describes 90% of the variance for the combined sets of orally bioavailable drugs in DrugBank, PROTACs in the PROTAC-DB 2.0[(p24)] (divided into CRBN, VHL and Other) and the 55 orally bioavailable PROTACs (CRBN, VHL and XIAP). The PCA was based on the descriptors of Lipinski's rule of 5,[(p4)] Veber's rule,[(p5)] plus the number of carbon atoms (nC) and the Kier flexibility index (Φ, Phi).[(p25)] The contribution of the descriptors to the PCA is indicated by arrows. The color coding of the compound sets in panels c and d is indicated in panel d. (d) Ellipses show the 95% confidence intervals for the orally bioavailable drugs in DrugBank, the three classes of PROTACs in the PROTAC-DB 2.0 and the 47 orally bioavailable CRBN PROTACs. The seven orally bioavailable VHL and the single XIAP PROTACs are indicated as filled circles. Abbreviations: AR: androgen receptor, ER: estrogen receptor.

have MWs ranging from just over 700 to close to 900 Da, whereas the two VHL PROTACs (XL01126 and ACBI2) that have bioavailabilities of 15% and 22% have MWs above 1,000 Da. We note that these MWs are close to or just beyond the guideline (MW ≤1,000 Da) deduced from a comprehensive analysis of oral drugs in the bRo5 chemical space.[(p8)]

In summary, among the orally bioavailable drugs from Drug-Bank, the oral CRBN and the oral VHL PROTACs are found in three different parts of chemical space defined by 2D descriptors, instead of in one overlapping oral chemical space (Figure 1d).

This observation illustrates a limitation of 2D descriptors for defining oral druggable space and points to the need for alternative strategies that allow the identification of PROTACs that are likely to be orally available.

Transitioning from 2D to 3D descriptors to improve the description of chemical space

As pointed out above, cell permeability is one of three crucial properties for oral bioavailability. A retrospective analysis of structure-permeability relationships for >35,000 compounds at

Pfizer suggested that 3D descriptors calculated from the predicted minimum energy conformations (MECs) of the compounds were more relevant than 2D ones, in particular for compounds residing in the bRo5 space.[p27] This observation originates from the structural complexity of compounds in the bRo5 space, which adopt distinct conformations that dictate their properties and that allow some to behave as molecular chameleons.[p28],[p29],[p30] The radius of gyration (R_{gyr}) was concluded to be a better descriptor of size than MW, and the 3D polar surface area (3D PSA) described polarity better than the HBA and HBD counts for these compounds.[p27] It is also widely accepted that the 3D structure and conformation dependent formation of intramolecular hydrogen bonds (IMHBs) often has a role in the absorption, distribution, metabolism and excretion (ADME) properties of drug candidates.[p31],[p32]

Below, we discuss recent results suggesting that the use of 3D descriptors provides advantages in the design of PROTACs. We also point to the value of using experimental physicochemical descriptors determined by high-throughput chromatography. In addition, we highlight some recent studies demonstrating that 3D descriptors calculated *in silico* correlate to experimental data and thereby have the potential to be used for the prediction of the physicochemical properties of PROTACs.

3D descriptors from NMR-derived conformations

Atazanavir, asunaprevir, daclatasvir, simeprevir, rifampicin, roxithromycin, telithromycin and spiramycin are cell permeable and orally bioavailable antiviral and antibacterial drugs that reside in the bRo5 chemical space. The conformational ensembles of these drugs that are likely to mediate cell permeability have been investigated by nuclear magnetic resonance (NMR) spectroscopy using chloroform ($\varepsilon = 4.8$) as a mimic of the interior of a cell membrane[p33] ($\varepsilon = 3.0$).[p34],[p35] It was found that the eight drugs populate a chemical space for which the population weighted mean values of the two 3D descriptors, R_{gyr} and solvent accessible 3D PSA, vary by up to 2.5 Å and 100 Å^2, respectively (Figure 2a).[p34],[p35] We hypothesize that this set of eight conformational ensembles describe a 3D chemical space in which it is also likely to find other orally bioavailable bRo5 chemical modalities.

In agreement with the above hypothesis, recent studies have revealed that conformational ensembles of cell-permeable CRBN and VHL PROTACs might populate the same 3D chemical space as the eight oral drugs (Figure 2a).[p36],[p37] The oral bioavailability of CRBN PROTACs **1–3** and VHL PROTAC **4** in Figure 2a have not been determined, but PROTACS **1** and **4** have high cell permeability, whereas **2** is moderately permeable and **3** has low permeability.[p36],[p37] NMR studies in chloroform revealed that the highly permeable PROTACs **1** and **4** had population weighted mean values of solvent accessible 3D PSA at the higher end of the range of the eight aforementioned drugs, whereas their R_{gyr} was at the low end (Figure 2a).[p36],[p37] The moderately permeable PROTAC **2** had a higher solvent accessible 3D PSA than PROTACs **1** and **4** and the eight drugs, but the conformations of lowly permeable PROTAC **3** could not be determined by NMR spectroscopy.[p37] However, the lack of long-range nuclear Overhauser effects suggested that PROTAC **3** populates extended conformations, having an even higher solvent accessible 3D PSA

and R_{gyr} than PROTAC **2**. In addition, PROTACs **1** and **4** adopted more polar and extended conformations in polar, protic environments than in chloroform: that is, they behaved as molecular chameleons (Figure 2b). This chameleonicity was concluded to contribute to the high cell permeability of PROTACs **1** and **4**, and was traced to their formation of intramolecular interactions, such as hydrogen bonds, which minimized size and polarity in non-polar environments.[p36],[p37]

In summary, these NMR studies suggest that 3D descriptors allow a better definition of cell-permeable PROTACs and the oral druggable space for PROTACs and bRo5 drugs than do 2D descriptors. However, the use of NMR spectroscopy for the determination of the solution ensembles, and thereby of 3D descriptors, currently requires weeks to months per compound, which prevents wider use in drug discovery projects.

Physicochemical descriptors based on high-throughput chromatographic methods

A recent literature survey revealed that PROTACs are rarely characterized with regards to their physiochemical properties, solubility and cell permeability.[p38] However, the determination of physicochemical descriptors related to solubility and permeability by chromatographic methods provides the opportunity to assess such properties with high-throughput and low consumption of material, thereby enhancing the efficiency of the PROTAC drug discovery pipeline.[p39] For instance, BRlogD[p40] characterizes the lipophilicity of a compound, similar to the octanol/water system; $\Delta logk_W^{IAM}$[p41] is an index of its polarity; and Chamelogk[p39] describes its chameleonicity. A compound with BRlogD >5 is considered excessively lipophilic and thus poorly soluble,[p42] whereas one with $\Delta logk_W^{IAM}$ >1.5 is overly polar; both lead to an unsatisfactory solubility/permeability profile. A chameleonic behavior could serve as a compensatory mechanism for these undesirable properties. Thus, compounds with Chamelogk >0.6 might exhibit oral bioavailability despite exceeding the limits of BRlogD and $\Delta logk_W^{IAM}$.

ARV-825 is a cell permeable and oral bioavailable CRBN-based PROTAC designed to degrade the bromodomain-containing protein 4 (BRD4) (Figure 2c).[p43] It exhibits an intermediate BRlogD and a reasonably low polarity. Additionally, it demonstrates an intermediate chameleonicity. ARV-825 thus possesses a lipophilicity/polarity profile that is suitable for oral administration, which can be further enhanced by its chameleonicity. This profile is significantly different from that of CMP98,[p44] a non-oral PROTAC. CMP98 has low lipophilicity, is highly polar and has a very low passive permeability in the parallel artificial membrane permeability assay (PAMPA, 2×10^{-9} cm/s).[p45] Although it shows chameleonic properties, this is not sufficient to overcome the high polarity and make CMP98 cell permeable and orally bioavailable. Overall, we propose that the three physicochemical descriptors, BRlogD, $\Delta logk_W^{IAM}$ and Chamelogk, can provide a rapid experimental characterization that indicates the likelihood of a PROTAC being orally bioavailable.

3D descriptors from *in silico* conformations

Ideally, 3D descriptors such as R_{gyr} and 3D PSA, calculated from *in silico* conformational ensembles, should be used to select cell-

KEYNOTE (GREEN)

KEYNOTE (GREEN)
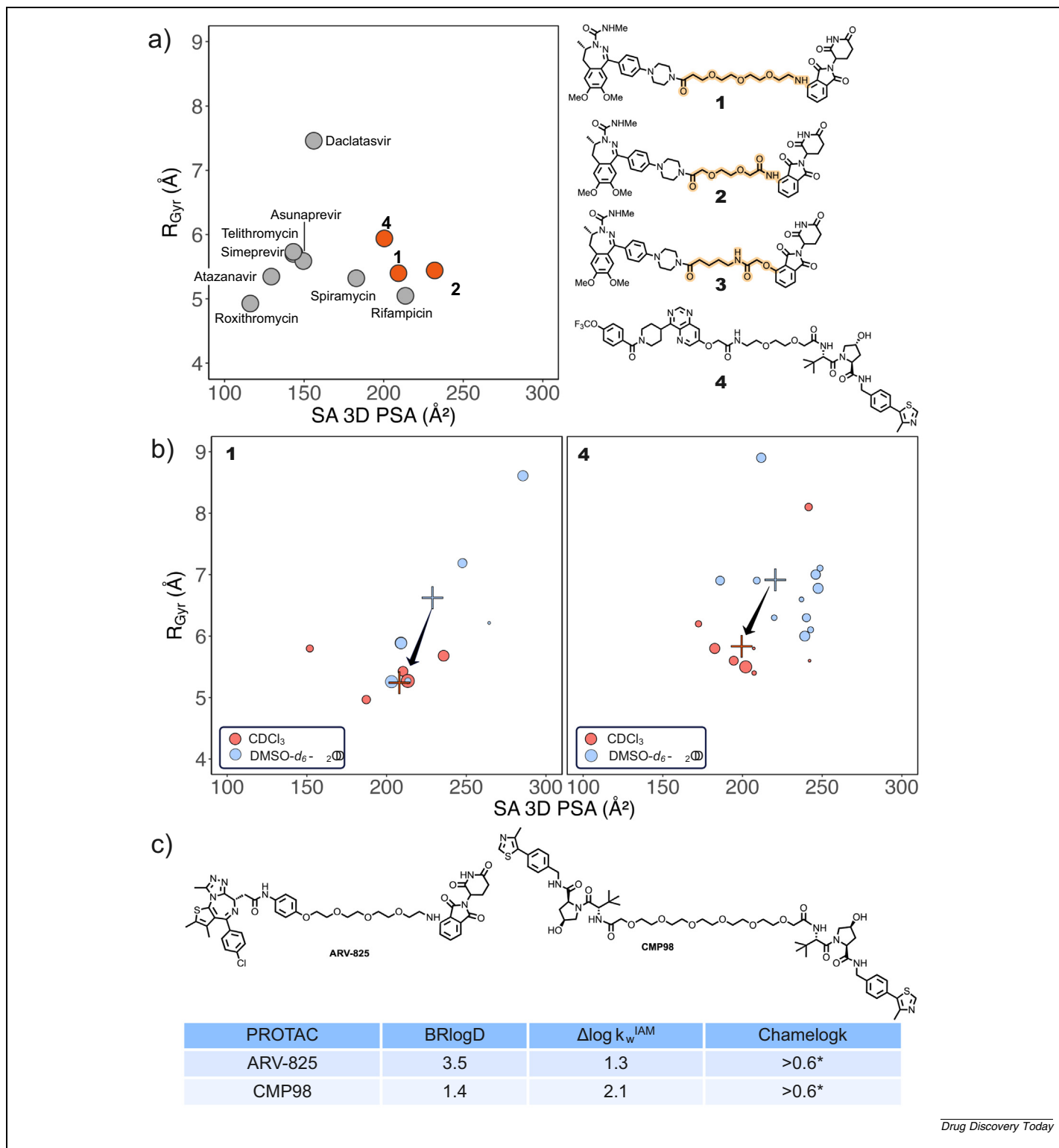
FIGURE 2

(a) Solvent accessible 3D polar surface area (SA 3D PSA) versus the radius of gyration (R_{gyr}) for eight orally bioavailable antiviral and antibacterial drugs (in gray) compared to those of PROTACs **1**, **2** and **4** (in red). Population weighted mean values of the two descriptors, calculated from the solution ensembles determined by ^{1}H NMR spectroscopy in CDCl$_3$ using the NAMFIS approach, are displayed for each compound. Alongside are the chemical structures for PROTACs **1–4**. As indicated in orange, **1–3** differ only in the structure of their linkers. **(b)** R_{gyr} versus SA 3D PSA for the conformations populated by the chameleonic PROTACs **1** and **4** in chloroform (red) and DMSO-water (blue), as determined by NMR spectroscopy. The area of each circle is proportional to the population of the corresponding conformation (in percentages). Population weighted mean values of the two descriptors are indicated by '+' signs in the color of the solvent used to determine the ensemble. The black arrows highlight that PROTACs **1** and **4** adopt more folded and less polar conformational ensembles in chloroform than in DMSO-water. **(c)** Structures and chromatographic descriptors of lipophilicity (BRlogD), polarity (Δlogk_W^{IAM}) and molecular chameleonicity (Chamelogk) for ARV-825 and CMP98. *According to Chamelogk definition.

permeable and soluble PROTACs for synthesis in the design process. As summarized below, in silico studies of the six PROTACs which have been studied experimentally by NMR spectroscopy (PROTACs 1–4 in Figure 2a) or chromatography (ARV-825 and CMP98 in Figure 2c) indeed indicate that this approach might be feasible.

For the CRBN PROTACs 1–3 (Figure 2a), which differ in permeability, molecular dynamics (MD) simulations were performed in explicit chloroform and were found to agree well with the NMR studies of these PROTACs.[p37] The simulations revealed that the propensity of the PROTACs to adopt folded and semi-folded conformations with low solvent accessible 3D PSA correlated to higher cell permeability, and that the shapes of the more frequently sampled conformations resembled those determined by NMR spectroscopy. The length, chemical nature and flexibility of the linker was concluded to be essential for allowing intramolecular interactions to stabilize folded conformations, having low solvent accessible 3D PSA for highly permeable PROTAC 1 in a non-polar environment. By contrast, the lowly permeable PROTAC 3 predominantly populated more extended and polar conformations. Conformational sampling for VHL PROTAC 4 was investigated using two different algorithms and the OPLS3e force field in implicit chloroform and water.[p46] Use of the Monte Carlo torsional sampling algorithm, followed by clustering and Boltzmann energy weighting of the conformations, provided in silico ensembles that overlapped well with those determined by NMR spectroscopy (Figure 3a). Thus, the sampled ensemble reproduced the chameleonic behavior of PROTAC 4, which was found to adopt conformations with lower solvent accessible 3D PSA in chloroform than in dimethyl sulfoxide (DMSO)-water by NMR spectroscopy. The difference between the sampled and experimental ensembles in a polar environment was assumed to originate from the fact that the sampled ensemble was generated in water, whereas the experimental one was determined in a mixture of DMSO and water (10:1) because of the low aqueous solubility of PROTAC 4.

In a recent article, three different computational approaches were used to compare the conformational spaces of the non-cell-permeable PROTAC CMP98[p45] and saquinavir, an oral HIV protease inhibitor.[p47] In silico ensembles were generated by (i) conformational sampling with implicit solvation, (ii) unbiased molecular dynamics and (iii) steered molecular dynamics (SMD); the latter two of which employed explicit solvation in a non-polar (toluene) and a polar (water) environment.[p47] All three methods found CMP98 to have a high molecular 3D PSA even in a nonpolar environment (Figure 3b, top), in agreement with the high polarity and low lipophilicity determined by chromatography (cf. Figure 2c). CMP98 was predicted to be chameleonic, which also agrees with its Chamelogk value (Figure 2c). However, as stated above, this chameleonicity is not sufficient to overcome the high polarity and make CMP98 cell permeable and orally bioavailable.

To allow a comparison between PROTACs, the SMD procedure used for the non-cell-permeable CMP98 was applied here to the orally bioavailable ARV-825[p43] (Figure 3b, bottom; methods in the supplementary material). As revealed by the molecular 3D PSA of in silico ensembles, ARV-825 was predicted to be significantly less polar than CMP98, just as observed by chromatography (Figure 2c). Furthermore, SMD found that ARV-825 adopted a less polar conformational ensemble in toluene than in water: that is, it behaved as a molecular chameleon in agreement its Chamelogk value (Figure 2c). The solvent accessible 3D PSA, which has been reported to correlate well to cell permeability,[p29] placed the toluene ensemble of ARV-825 within the chemical space populated by oral drugs, whereas that of CMP98 was located outside the oral space (Figure S2 in the supplementary material; Figure 2a). Altogether these predictions agree well with ARV-825 being cell permeable and orally bioavailable, CMP98 displays neither property.

For oral bRo5 drugs and cell-permeable PROTACs, permeability has been found to be favored for compounds that are able to adopt ensembles of well-defined, folded conformations with low surface accessible 3D PSA in a non-polar environment.[p29],[p37] Notably, there are striking differences in the shape (R_{gyr}) of the conformational ensemble of the two PROTACs (Figure 3b). ARV-825 displays a large variety of folded and extended polar conformations in water, whereas in a non-polar environment it only exists in folded conformations. On the contrary, CMP98 displays the inverse pattern and populates a large diversity of conformations in a non-polar environment, whereas conformations in water are highly folded. Consequently, the differences in folding in a non-polar environment predicted by SMD also indicate that ARV-825 is the more permeable of the two PROTACs.

In summary, in silico techniques, including Monte Carlo conformational sampling, unbiased MD and steered MD simulations, provide conformational ensembles from which 3D descriptors can be calculated that agree reasonably well with the experimentally determined ones. Generating conformational ensembles for predicting 3D descriptors is a more time-consuming process than calculating 2D descriptors, but because of their enhanced informativeness, 3D descriptors are more appealing for application in the quest to discover orally bioavailable PROTACs. Because only a few PROTACs have been studied by in silico techniques, studies of larger compound sets are required to determine their scope and limitations. It also remains to be established whether certain combinations of algorithms and forcefields are more applicable to PROTACs than others, along with determining the best approach for identifying the biologically relevant conformations. The development of algorithms and forcefields that allow fast and accurate prediction of intramolecular interactions formed upon transitioning from an aqueous to a membrane-like environment should improve the reliability of cell permeability predictions in the future.

Considerations for design

Current strategies for delivery of orally bioavailable PROTACs often involve the optimization of the molecular descriptors of Lipinski's and Veber's rule of the three parts of the PROTACs, and thereby also of the overall compound[p14],[p15],[p16]: that is, the strategy used for Ro5-compliant small molecule drugs. To aid PROTAC drug discovery, outer descriptor limits for the oral druggable space of PROTACs have been derived by analysis of more than 1,800 proprietary PROTACs,[p16] and from a survey of 18 pharmaceutical companies (Figure 4a).[p48] The majority
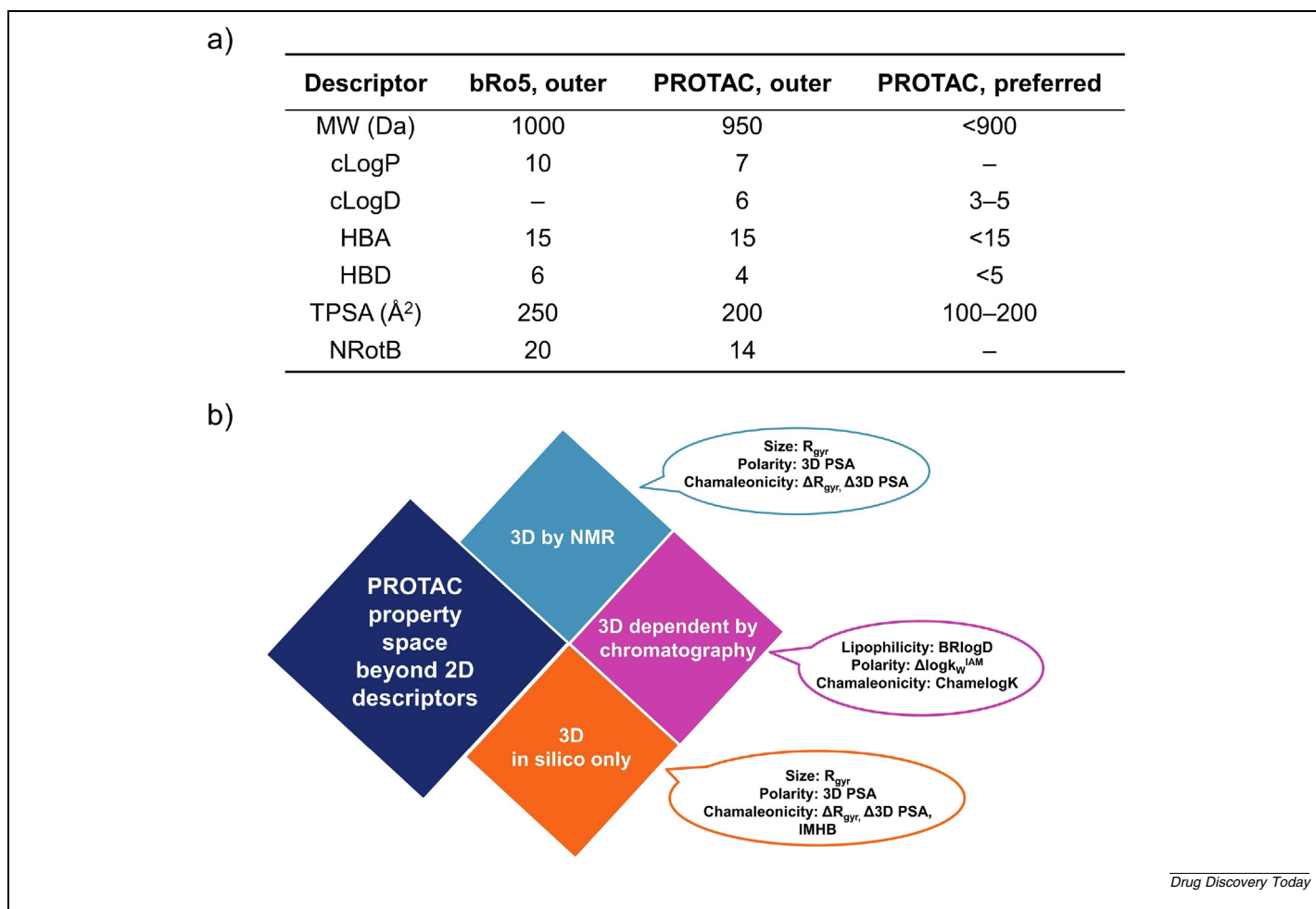
KEYNOTE (GREEN)
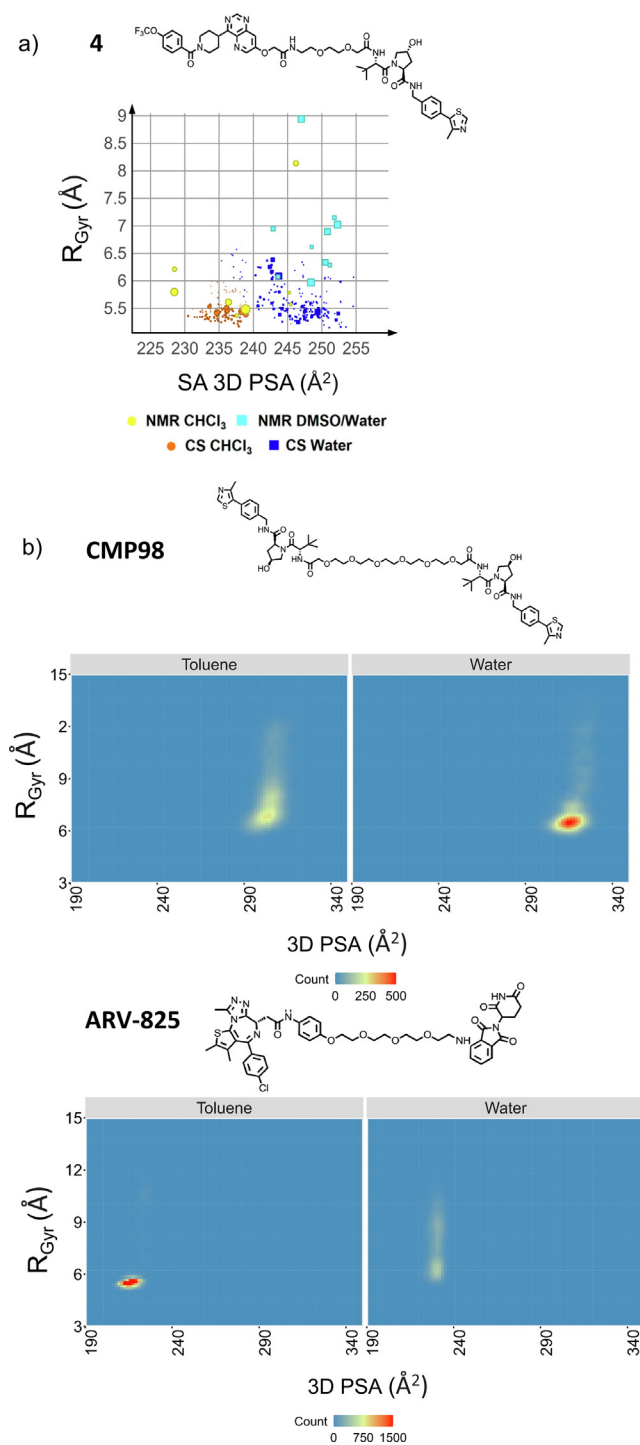
FIGURE 3

(a) Density plot of R_{gyr} versus solvent accessible 3D PSA of the *in silico* ensembles from conformational sampling (CS) of PROTAC 4 compared to those determined experimentally by NMR spectroscopy. (b) Density plot of R_{gyr} versus molecular 3D PSA for the conformational ensembles of CMP98 and ARV-825 obtained from steered moleculur dynamics (SMD) simulations.

of the PROTACs that have an oral bioavailability >5% in the data set reported herein are located within these limits, with the two VHL-based PROTACs being outliers (cf. the data set in the supplementary material). As might have been expected, these guidelines agree fairly well with those of the bRo5 space established by studies of oral drugs, clinical candidates and

a)

Descriptor	bRo5, outer	PROTAC, outer	PROTAC, preferred
MW (Da)	1000	950	<900
cLogP	10	7	–
cLogD	–	6	3–5
HBA	15	15	<15
HBD	6	4	<5
TPSA (Å^2)	250	200	100–200
NRotB	20	14	–

b)

FIGURE 4

(a) Comparison of the outer limits of oral druggable space for drugs and clinical candidates in the bRo5 space[p8] and those derived from PROTACs.[p16],[p48] The upper limits of the 2D descriptor ranges for oral PROTACs were obtained by analysis of more than 1,800 proprietary PROTACs at Arvinas (PROTAC, outer),[p16] and from a survey of 18 pharmaceutical companies (PROTAC, preferred).[p48] (b) Overview of the toolbox of techniques and the experimental and *in silico* 3D descriptors that can be used for the identification and optimization of orally bioavailable PROTACs.

compounds in advanced stages of lead optimization.[p8],[p9] Discovery of cell-permeable and orally bioavailable PROTACs close to, or beyond, the limits of this chemical space is very challenging and will most likely require judicious optimization, for example by 3D approaches that include the design of molecular chameleons.

Models built for prediction of the aqueous solubility and cell permeability of PROTACs also provide some guidance for the design of oral PROTACs. A random tree model for the thermodynamic aqueous solubility was recently built using a small set of 15 CRBN and VHL PROTACs as training set.[p42] According to the model, PROTACs that have a TPSA >289 Å^2 have a high probability of being highly soluble (>200 μM). Less polar PROTACs can be classified into low (<30 μM) and intermediate (30–200 μM) solubility groups by determination of the chromatographic lipophilicity descriptor BRlogD. By contrast, random forest classification models built for the identification of VHL PROTACs with high or low cell permeability using a training set of 253 VHL PROTACs revealed a more complex correlation to individual descriptors.[p49] All 17 descriptors, which included those of the Ro5 and Veber's rule, were important for the models, with those of lipophilicity and molecular size having only two- to threefold higher weight. It is possible that a high degree of chameleonicity in this VHL PROTAC data set explains the high dependency of the models on all descriptors.

The linker is the moiety, which has the largest potential for variation in the design of PROTACs with optimized pharmacokinetic properties. All but one of the 33 CRBN PROTACs in our data set that have a bioavailability ≥5% have linkers that contain at least one heterocyclic moiety, mainly piperidines and/or piperazines (Figure S3 in the supplementary material). By contrast, three of the four bioavailable (F >4%) VHL PROTACs have linkers based on alkyl chains (Figure S4 in the supplementary material). This gives the impression that orally bioavailable CRBN PROTACs are much more rigid than the VHL PROTACs, and potentially less able to behave as molecular chameleons. To the best of our knowledge, the flexibility and chameleonicity of PROTACs containing heterocyclic linkers has not yet been investigated. However, piperidine and piperazine moieties retain significant conformational flexibility, as revealed by the Kier flex-

ibility index (Φ)[(p25),(p26)] of the CRBN PROTACs (median Φ of 11, Figure S5 in the supplementary material). Three of the four VHL PROTACs have Φ 14–15, while the fourth has Φ 17 (Figure S5 in the supplementary material). In comparison, four bRo5 drugs found to behave as molecular chameleons using NMR spectroscopy,[(p35)] by determination of Chamelogk[(p39)] or by analysis of crystal structures[(p30)] have similar flexibilities, with Φ ranging from 11 to 13. PROTACs **1** and **4** in Figure 2a have Φ 16 and 19, respectively. In conclusion, CRBN PROTACs based on heterocyclic linkers, just like VHL PROTACs with linkers based on alkyl or ethylene glycol chains, have a flexibility sufficient for allowing them to behave as molecular chameleons. We therefore propose that prediction of *in silico* conformations and chromatographic determination of Chamelogk are approaches that should be considered in the design of PROTACs with optimal pharmacokinetic properties.

Concluding remarks

Large and flexible compounds, such as PROTACs, reside far into the bRo5 chemical space, revealing that they are at high risk of not becoming oral drugs. Recent studies, however, show that for some, but not all, this risk can be overcome. Unsurprisingly, CRBN PROTACs, which are found closer to the chemical space defined by the 2D descriptors of Lipinski's and Veber's rules, carry less risk than PROTACs based on VHL and other complex E3 ligase ligands. This is well illustrated by the clinical pipeline of PROTACs and the data set of PROTACs that show oral bioavailability in rodents discussed herein (Figure 1a).[(p22),(p48),(p50)]

Some orally bioavailable VHL and high-molecular-weight CRBN PROTACs have already been discovered, and new design strategies that release the potential of such PROTACs are in high demand. In this minireview, we propose that 3D descriptors, such as those of size (R_{gyr}) and polarity (surface accessible 3D PSA), provide a better definition of the cell-permeable and orally druggable space of PROTACs (Figure 4b). This is because, in contrast to 2D descriptors, 3D descriptors describe the properties of large and stereochemically complex compounds, such as PROTACs and bRo5 drugs, in a uniform way. 3D descriptors also capture the compounds' ability to behave as molecular chameleons by adapting their structure and properties to the environment. We also highlight recent results that suggest that physicochemical properties determined by chromatographic methods, including the determination of chameleonicity, show potential for the identification and optimization of cell-permeable and orally absorbed PROTACs. In addition, results indicate that conformational ensembles generated *in silico* could be used to predict relevant 3D descriptors, and thereby could be used to design orally absorbed PROTACs. We suggest that the combined use of *in silico* and chromatographic methods is poised to improve the success rate of the discovery of oral PROTACs, and that the use of these techniques will be of particular importance for VHL PROTACs that reside at the outer borders of oral bRo5 space.

CRediT authorship contribution statement

Giulia Apprato: Conceptualization, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Vasanthanathan Poongavanam:** Conceptualization, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Diego Garcia Jimenez:** Data curation, Visualization, Writing – original draft. **Yoseph Atilaw:** Formal analysis, Writing – original draft. **Mate Erdelyi:** Formal analysis, Writing – review & editing. **Giuseppe Ermondi:** Data curation, Methodology. **Giulia Caron:** Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing. **Jan Kihlberg:** Conceptualization, Formal analysis, Funding acquisition, Visualization, Writing – original draft, Writing – review & editing.

Data availability

Data is shared in the Supporting Information

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.drudis.2024.103917.

References

1. Liu X, Ciulli A. Proximity-based modalities for biology and medicine. *ACS Cent Sci*. 2023;9:1269–1284.
2. Békés M et al. PROTAC targeted protein degraders: the past is prologue. *Nat Rev Drug Disc*. 2022;21:181–200.
3. Weng G et al. PROTAC-DB: an online database of PROTACs. *Nucleic Acids Res*. 2021;49:D1381–D1387.
4. Lipinski CA et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 1997;23:3–25.
5. Veber DF et al. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem*. 2002;45:2615–2623.
6. Edmondson SD et al. Proteolysis targeting chimeras (PROTACs) in 'beyond rule-of-five' chemical space: recent progress and future challenges. *Bioorg Med Chem Lett*. 2019;29:1555–1564.
7. Maple HJ et al. Developing degraders: principles and perspectives on design and chemical space. *Med Chem Commun*. 2019;10:1755–1764.
8. Doak BC et al. Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates. *Chem Biol*. 2014;21:1115–1142.
9. DeGoey DA et al. Beyond the rule of 5: lessons learned from AbbVie's drugs and compound collection. *J Med Chem*. 2018;61:2636–2651.
10. Poongavanam V, Kihlberg J. PROTAC cell permeability and oral bioavailability: a journey into uncharted territory. *Future Med Chem*. 2021;14:123–126.

11. O'Brien Laramy MN et al. Delivering on the promise of protein degraders. *Nat Rev Drug Disc*. 2023;22:410–427.

12. Sastry SV et al. Recent technological advances in oral drug delivery – a review. *Pharm Sci Technol Today*. 2000;3:138–145.

13. Aungst BJ. Optimizing oral bioavailability in drug discovery: an overview of design and testing strategies and formulation options. *J Pharm Sci*. 2017;106:921–929.

14. Han X, Sun Y. Strategies for the discovery of oral PROTAC degraders aimed at cancer therapy. *Cell Rep Phys Sci*. 2022;3 101062.

15. Zeng S et al. Current advances and development strategies of orally bioavailable PROTACs. *Eur J Med Chem*. 2023;261 115793.

16. Hornberger KR, Araujo EMV. Physicochemical property determinants of oral absorption for PROTAC protein degraders. *J Med Chem*. 2023;66:8281–8287.

17. Bondeson DP et al. Catalytic *in vivo* protein knockdown by small-molecule PROTACs. *Nat Chem Biol*. 2015;11:611–617.

18. Petrylak DP et al. A phase 2 expansion study of ARV-766, a PROTAC androgen receptor (AR) degrader, in metastatic castration-resistant prostate cancer (mCRPC). *J Clin Oncol*. 2023;41:TPS290.

19. Gao X et al. Phase 1/2 study of ARV-110, an androgen receptor (AR) PROTAC degrader, in metastatic castration-resistant prostate cancer (mCRPC). *J Clin Oncol*. 2022;40:017.

20. Campone M et al. VERITAC-2: A global, randomized phase 3 study of ARV-471, a proteolysis targeting chimera (PROTAC) estrogen receptor (ER) degrader, vs fulvestrant in ER+/human epidermal growth factor receptor 2 (HER2)- advanced breast cancer. *J Clin Oncol*. 2023;41:TPS1122.

21. Kofink C et al. A selective and orally bioavailable VHL-recruiting PROTAC achieves SMARCA2 degradation *in vivo*. *Nat Comm*. 2022;13:5969.

22. Pike A et al. Optimising proteolysis-targeting chimeras (PROTACs) for oral drug delivery: a drug metabolism and pharmacokinetics perspective. *Drug Discov Today*. 2020;25:1793–1800.

23. Wishart DS et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46:D1074–D1082.

24. Weng G et al. PROTAC-DB 2.0: an updated database of PROTACs. *Nucleic Acids Res*. 2022;51:D1367–D1372.

25. Caron G et al. Flexibility in early drug discovery: focus on the beyond-Rule-of-5 chemical space. *Drug Discov Today*. 2020;25:621–627.

26. Kier LB. An index of molecular flexibility from kappa shape attributes. *Mol Inform*. 1989;8:221–224.

27. Guimarães CRW et al. Use of 3D properties to characterize beyond rule-of-5 property space for passive permeation. *J Chem Inf Model*. 2012;52:882–890.

28. Whitty A et al. Quantifying the chameleonic properties of macrocycles and other high-molecular-weight drugs. *Drug Discov Today*. 2016;21:712–717.

29. Rossi Sebastiano M et al. Impact of dynamically exposed polarity on permeability and solubility of chameleonic drugs beyond the rule of 5. *J Med Chem*. 2018;61:4189–4202.

30. Poongavanam V et al. Molecular chameleons in drug discovery. *Nat Rev Chem*. 2024;8:45–60.

31. Kuhn B et al. Intramolecular hydrogen bonding in medicinal chemistry. *J Med Chem*. 2010;53:2601–2611.

32. Caron G et al. Intramolecular hydrogen bonding: an opportunity for improved design in medicinal chemistry. *Med Res Rev*. 2019;39:1707–1729.

33. Gramse G et al. Nanoscale measurement of the dielectric constant of supported lipid bilayers in aqueous solutions with electrostatic force microscopy. *Biophys J*. 2013;104:1257–1262.

34. Danelius E et al. Solution conformations explain the chameleonic behavior of macrocyclic drugs. *Chem Eur J*. 2020;26:5231–5244.

35. Wieske LHE et al. Going viral: an investigation into the chameleonic behaviour of antiviral compounds. *Chem Eur J*. 2023;29:e202202798.

36. Atilaw Y et al. Solution conformations shed light on PROTAC cell permeability. *ACS Med Chem Lett*. 2021;12:107–114.

37. Poongavanam V et al. Linker-dependent folding rationalizes PROTAC cell permeability. *J Med Chem*. 2022;65:13029–13040.

38. Apprato G et al. The quest for oral PROTAC drugs: evaluating the weaknesses of the screening pipeline. *ACS Med Chem Lett*. 2023;14:879–883.

39. García Jiménez D et al. Chamelogk: a chromatographic chameleonicity quantifier to design orally bioavailable beyond-rule-of-5 drugs. *J Med Chem*. 2023;66:10681–10693.

40. Ermondi G et al. Experimental lipophilicity for beyond Rule of 5 compounds. *Future Drug Discov*. 2019;1:2.

41. Ermondi G et al. Learning how to use IAM chromatography for predicting permeability. *Eur J Pharm Sci*. 2018;114:385–390.

42. García Jiménez D et al. Designing soluble PROTACs: strategies and preliminary guidelines. *J Med Chem*. 2022;65:12639–12649.

43. He L et al. ARV-825-induced BRD4 protein degradation as a therapy for thyroid carcinoma. *Aging*. 2020;12:4547–4557.

44. Maniaci C et al. Homo-PROTACs: bivalent small-molecule dimerizers of the VHL E3 ubiquitin ligase to induce self-degradation. *Nat Commun*. 2017;8:830.

45. Klein VG et al. Understanding and improving the membrane permeability of VH032-based PROTACs. *ACS Med Chem Lett*. 2020;11:1732–1738.

46. Ermondi G et al. Conformational sampling deciphers the chameleonic properties of a VHL-based degrader. *Pharmaceutics*. 2023;15:272.

47. Rossi Sebastiano M et al. Refinement of computational access to molecular physicochemical properties: From Ro5 to bRo5. *J Med Chem*. 2022;65:12068–12083.

48. Volak LP et al. Industry perspective on the pharmacokinetic and absorption, distribution, metabolism, and excretion characterization of heterobifunctional protein degraders. *Drug Metab Dispos*. 2023;51:792–803.

49. Poongavanam V et al. Predictive modeling of PROTAC cell permeability with machine learning. *ACS Omega*. 2023;8:5901–5916.

50. Cantrill C et al. Fundamental aspects of DMPK optimization of targeted protein degraders. *Drug Discov Today*. 2020;25:969–982.