Synergistic Target Combinations Against Obesity: Focus on MCHR1/H$_3$R Modulation

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by
David Andreas Schaller
from Berlin, Germany

2020

The presented thesis was prepared from November 2014 till March 2020 under the supervision of Prof. Dr. Gerhard Wolber at the Institute of Pharmacy of the Freie Universität Berlin.

Supervisor: Prof. Dr. Gerhard Wolber
Second examiner: Prof. Dr. Johannes Kirchmair

Date of Defense: June 29$^{\text{th}}$ 2020

Acknowledgements

Abstract

The continuing epidemic of overweight and obesity faces a lack of appropriate pharmaceutical treatment options to support health care systems. Although studied for decades, currently available drugs only show low efficacy but serious or at least unpleasant side effects. Multi-target drugs hold promise to overcome the limitations of traditional pharmaceutics by modulating several nodes of the disease-relevant biological network. In this thesis, the multi-target concept was applied to macromolecular targets involved in obesity.

In order to identify target pairs useful as starting points for multi-target drug design, we applied a systematic data mining approach employing publicly available bioactivity data of small molecules binding to targets involved in obesity. The target pair with the highest molecular similarity among known active ligands was found to comprise of histamine H_3 receptor (H_3R) and melanin-concentrating hormone receptor 1 (MCHR1). Both proteins are part of the G-protein coupled receptor (GPCR) family and were extensively studied as potential obesity targets. Although antagonizing either receptor was efficient in rodent models of obesity, drug candidates failed to proof efficacy in clinical studies. To test the potential of H_3R and MCHR1 in multi-target drug development, a shape-based virtual screening campaign was conducted resulting in the selection of three small molecules. A subsequent *in vitro* evaluation revealed nanomolar affinity for all three molecules at both receptors.

Lead optimization against multiple targets can dramatically benefit from integration of structural data. Since H_3R and MCHR1 lack experimental structural data for structure-based drug design, two novel methods were developed that support drug design campaigns based on homology models. H_3R is part of the aminergic family of GPCRs, which share a conserved charged interaction between ligand and protein. This crucial interaction was incorporated into a ligand-guided homology modeling campaign revealing valuable insights into side chain conformations critical for appropriate ligand placement in H_3R. A subsequent virtual screening campaign followed by *in vitro* validation revealed two novel ligands with nanomolar affinity at H_3R. MCHR1 is less well characterized and was found to contain several highly flexible residues in the ligand binding pocket, which hindered the translation of the ligand-guided homology modeling strategy to MCHR1. To include the high flexibility of binding site residues, the protein environment of water molecules in molecular dynamics simulations was analyzed to derive 3D pharmacophores for virtual screening. Generated 3D pharmacophores were highly successful in a retrospective virtual screening campaign in discriminating active MCHR1 ligands from decoys. This method was translated into a Python package (PyRod), where the source code was released publicly.

The results and methods developed in this thesis provide valuable tools to support the development of more efficient and safe anti-obesity medications. We show that

the simultaneous inhibition of H_3R and MCHR1 with a single high affinity binder is possible. We developed novel computational methods to support structure-based virtual screening campaigns against both receptors.

Zusammenfassung

Die ständig ansteigende Prävalenz für Übergewicht und Adipositas offenbart einen dramatischen Mangel an geeigenten pharmazeutischen Behandlungsoptionen. Obwohl seit Jahrzenten an einer Behandlung geforscht wird, weisen derzeitig verfügbare Pharmazeutika nur eine geringe Wirksamkeit und schwere oder zumindest unangenehme Nebenwirkungen auf. Wirkstoffe, welche die Aktivität mehrerer Zielstrukturen modulieren können, versprechen die Limitationen traditioneller Medikamente zu überwinden, indem sie das biologische Netzwerk hinter einer Krankheit an verschiedenen Stellen simultan beeinflussen. In dieser Dissertation wurde das Konzept zur Modulierung mehrerer Zielstrukturen auf Adipositas angewendet.

Um Adipositas-relevante Zielstruktur-Paare zu identifizieren, deren Aktivität mit einem Molekül moduliert werden könnte, wurde ein systematischer Ansatz zur Analyse von öffentlich zugänglichen Aktivitätsdaten kleiner Moleküle durchgeführt. Das Zielstruktur-Paar mit der größten Ähnlichkeit zwischen bekannten Liganden beinhaltet den Histamine H_3 Rezeptor (H_3R) und den Rezeptor 1 des Melanin-konzentrierenden Hormons (MCHR1). Beide Proteine sind Teil der G-Protein gekoppelten Rezeptor Familie und wurden intensiv für die Behandlung von Adipositas untersucht. Obwohl Antagonisten für beide Rezeptoren bei Nagetier-Modellen für Adipositas wirksam waren, konnten diese vielversprechenden Ergebnisse nicht auf den Menschen übertragen werden. Um das Potential von H_3R und MCHR1 als Zielstruktur-Paar zu überprüfen, wurde eine 3D Ähnlichkeitssuche durchgeführt, welche in der Identifizierung von drei Molekülen resultierte. Eine anschließende *in vitro* Validierung zeigte nanomolare Affinitäten für alle drei Moleküle an beiden Rezeptoren.

Die Optimierung von Wirkstoffkandidaten gegen mehrere Zielstrukturen kann erheblich von der Integerierung von Strukturdaten profitieren. Da für H_3R und MCHR1 keine experimentel bestätigten Strukturen zur Verfügung standen, wurden zwei neue Methoden entwickelt, welche auf Homologie Modellen-basierende Projekte zur Wirkstoffentwicklung unterstützen können. H_3R ist Teil der aminergen Familie von GPCRs, welche eine konservierte geladene Interaktion zwischen Ligand und Protein aufweisen. Diese entscheidende Interaktion wurde in einem Liganden-geleiteten Homologie Modelierungs-Ansatz ausgenutzt und ermöglichte die Charakterisierung von Seitenketten-Konformationen, die eine korrekte Platzierung der Liganden ermöglichen. Ein virtuelles Screening gefolgt von einer *in vitro* Validierung identifizierte zwei neue Moleküle mit nanomolarer Affinität am H_3R. MCHR1 wurde weniger stark untersucht und beinhaltet mehrere hoch flexible Seitenketten in der Bindetasche, welche eine Übertragung des Liganden-geleiteten Homologie Modelierungs-Ansatzes verhindern. Um die hohe Flexibilität der Bindetasche zu berücksichtigen, wurde die Protein-Umgebung von Wasser-Molekülen in Simulationen zur Moleküldynamik analysiert um 3D Pharmakophore für virtuelles Screening abzuleiten. Die generierten 3D Pharmakophore waren

in einem retrospektiven Screening in der Lage aktive MCHR1 Liganden von inaktiven Molekülen zu unterscheiden. Diese Methode wurde in ein Python Packet (PyRod) übertragen, dessen Quellcode öffentlich zugänglich gemacht wurde.

Die in dieser Arbeit entstandenen Ergebnisse und Methoden bieten wertvolle Werkzeuge für die Entwicklung von effektiven und sicheren Adipositas-Wirkstoffen. Wir konnten zeigen, dass die simultane Inhibition von H_3R und MCHR1 mit einem einzigen Molekül möglich ist. Wir entwickelten neue computergestütze Methoden, um strukturbasiertes virtuelles Screening für beide Rezeptoren zu unterstützen.

Contents

Introduction

1.1   The Obesity Epidemic

Obesity and overweight have evolved into significant threats of human health account-
ing for 4 Mio deaths in 2015 [1]. These medical conditions are characterized by an
abnormal or excessive accumulation of fat, and are associated with various serious
diseases including type 2 diabetes, hypertension, myocardial infarction, stroke, venous
thromboembolism and certain types of cancer [2]. Since 1980 the prevalence of obesity
has constantly increased affecting the health of 13 % of adults world-wide in 2016
(Fig 1) [3]. Although studied for decades, health care systems still lack an effective
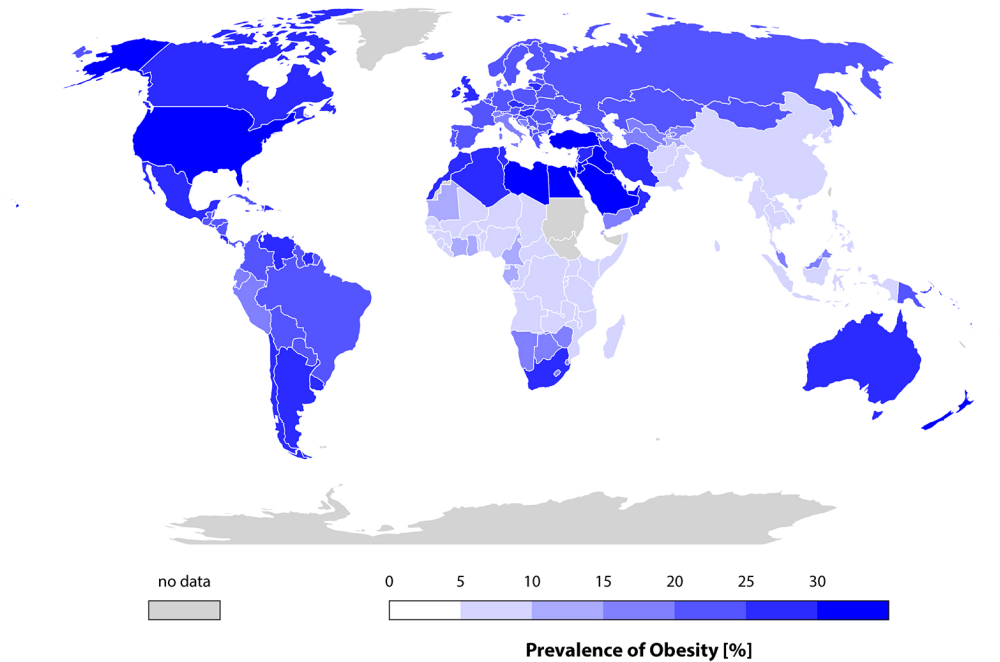treatment to stop the obesity epidemic.

Figure 1: Prevalence of obesity among adults. Data from World Health Organization [3].

1.2   Regulation of Energy Homeostasis

Main causes for overweight and obesity are genetic susceptibility, increased consumption of high-energy food and insufficient physical activity, resulting in the the excess of energy intake over energy expenditure [2].

Energy homeostasis is regulated by a multitude of hormones such as leptin, insulin and ghrelin mainly secreted by gut, pancreas and adipose tissue. These peripheral signals either act as appetite stimulant (orexigenic) or appetite suppressant (anorectic) by binding to their cognate receptors and thus stimulate the central nervous system (CNS), notably the arcuate nucleus (ARC) of the hypothalamus (Fig 2). Within the ARC, two types of neurons exist that either co-express the orexigenic neurotransmitters neuropeptide Y (NPY) and agouti-related Protein (AgRP) or co-express the anorectic neurotransmitters pro-opiomelanocortin (POMC) as well as cocaine and amphetamine-related transcript (CART) [4, 5].

The orexigenic NPY/AgRP neurons can inhibit anorectic POMC neurons by secreting γ-aminobutyric acid (GABA). Both types of neurons more widely project into the CNS including the paraventricular nucleus (PVN) and the lateral hypothalamic area (LHA). These areas home putative second-order neurons that also interact with other anorectic and orexigenic signals from gut and CNS mediating behavioral, endocrine and autonomous effects of changes in energy status by acting on neurons of the nucleus tractus solitaries (NTS) [4, 5].

1.3   Targets for Obesity Treatment

The major challenge for developing drugs against obesity is the complexity of metabolic body weight control. It is maintained by an equilibrium between three major pathways: food intake, energy generation and fat storage [6]. Consequently, this model provides different opportunities for pharmaceutical intervention: (i) agents that decrease appetite through central action, (ii) agents that affect metabolism through peripheral action, (iii) agents modulating the gastrointestinal (GI) tract and (iv) agents that not only affect obesity, but also reduce mortality and morbidity of associated disorders. Current research on obesity related pathways reveals a plethora of receptors, enzymes and transcription factors that might be useful as targets for anti-obesity agents (Tab 1). In the following sections, examples are provided highlighting several targets for drug development against obesity.

1.3.1   Decreasing Appetite Through Central Action

Leptin is a hormone expressed and secreted by adipocytes in proportion to body fat stores [33]. It acts anorectic through its interaction with the leptin receptor, a single membrane-spanning class I cytokine receptor with tyrosine kinase activity. Among
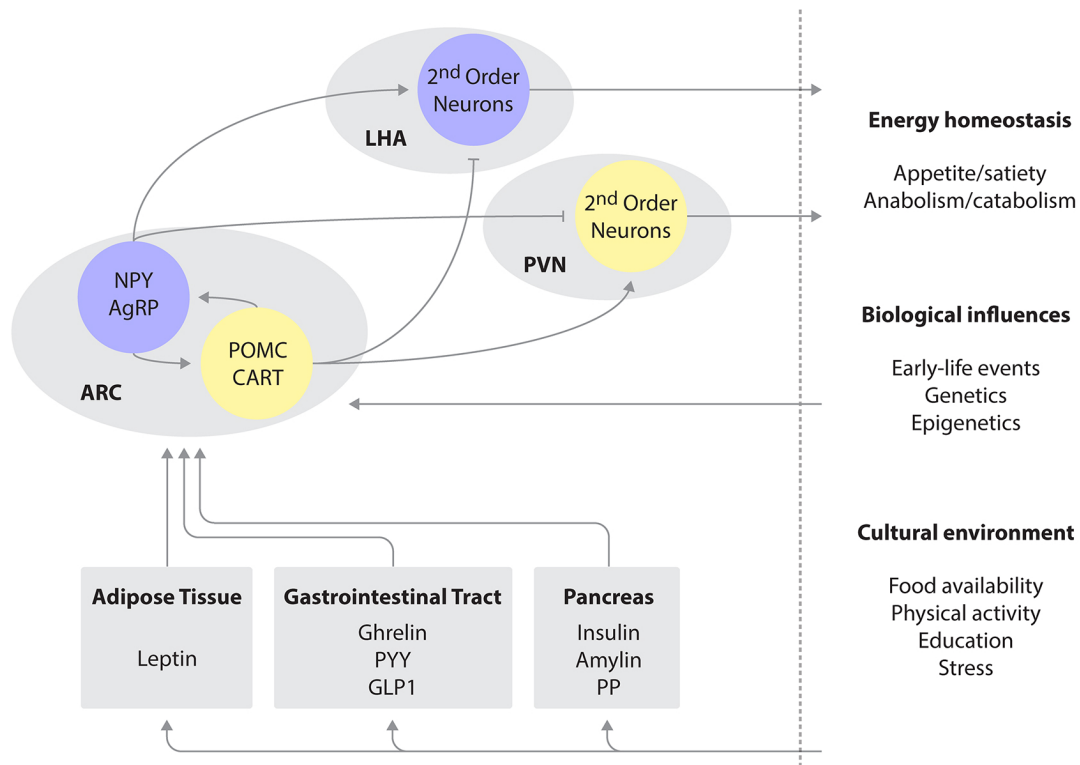
Figure 2: Regulation of energy homeostasis by neuronal circuits and extrinsic influences. AgRP - agouti-related Protein, ARC - arcuate nucleus, CART - cocaine and amphetamine-related transcript, GLP1 - Glucagon-like peptide 1, LHA - lateral hypothalamic area, NPY - neurotransmitters neuropeptide Y, POMC - pro-opiomelanocortin, PP - pncreatic polypeptide, PVN - paraventricular nucleus, PYY - peptide YY.

rare individuals who are obese because of the lack of this peptide, administration of physiological doses of leptin decreases food intake and causes weight loss. However, common obese patients are leptin-resistant and have elevated circulating levels of this peptide. Experimental studies in animals have identified two intracellular proteins that terminate receptor signaling, acting as potent mediators of leptin resistance: cytokine signaling- 3 (SOCS3) and protein tyrosine phosphatase 1B (PTP1B). Therefore, inhibition of either of these proteins could increase leptin sensitivity. However, inhibition of SOCS3 as therapeutic strategy for obesity treatment is highly advised because this protein is implicated in several biological processes [35]. In contrast, PTP1B is a viable anti-obesity drug target, since PTP1B knockout rats are sensitive to leptin and insulin and resistant to diet-induced obesity. Moreover, it has been shown that the selective blockade of the PTP1B expression results in decreases in food intake and reduction of body weight by increasing the action of leptin and insulin in the hypothalamus [36]. These results validate PTP1B inhibitors as promising anti-obesity agents.

Table 1: Obesity-relevant targets retrieved from a literature research covered in the systematic data mining workflow presented in section 4.1 [7]. Only targets were included, for which ligand data has been deposited in the ChEMBL database [8] to the time of the study.

Required Modulation	Receptors		Enzymes	Transcription factors
Activation	$5HT_{1B}R$ [9]	$5HT_{2C}R$ [9]	SIRT1 [10]	PPARα [11]
	AMY1 [12]	AMY3 [12]		PPARδ [11]
	$β_3AR$ [13]	BRS3 [14]		PPARγ [11]
	CCKAR [15]	GLP1R [15]		THRβ [16]
	MCR3 [17]	MCR4 [17]		
	NPYR2 [18]	NPYR4 [18]		
	OX1R [19]	OX2R [19]		
Deactivation	$5HT_6R$ [20]	CB1 [21]	11βHD1 [22]	
	CRHR2 [23]	GALR1 [24]	ACC1 [25]	
	GHSR [15]	H_3R [26]	ACC2 [25]	
	MCHR1 [27]	μOR [28]	CPT1L [29]	
	NPYR1 [18]	NPYR5 [18]	CPT1M [29]	
			DGAT1 [30]	
			FAS [31]	
			PLIP [32]	
			PTP1B [33]	
			SCD1 [34]	

1.3.2 Affecting Metabolism Through Peripheral Action

Activation of the $β_3$ adrenergic receptor ($β_3AR$) induces both catecholamine-stimulated lipolysis and thermogenesis in adipose tissue [13]. It appears that thermogenesis is primarily responsible for the removal of stored fat in animal models. Treatment with molecules selectively activating $β_3AR$ markedly increases energy expenditure and decreases obesity in rodents [37]. These observations demonstrate the potential of $β_3AR$ activators as anti-obesity agents.

1.3.3 Modulating the Gastrointestinal Tract

Cholecystokinin (CCK) inhibits food intake in all mammalian species in which it has been tested, including human [15]. This peptide is rapidly released from L cells in the upper intestine in response to the intraluminal presence of digestive products, resulting in the earlier termination of the meal. The effect of CCK on feeding is mediated by its interaction with the CCK A receptor (CCKAR), which is expressed

in the GI tract as well as in several areas of the CNS. Infusion of the C-terminal octapeptide of CCK, the shortest bioactive form, reduces food intake in obese men [38]. In contrast, administration of CCKAR inhibitors results in increased food intake through increasing meal size [39].

1.3.4 Reducing the Morbidity of Associated Disorders

The peroxisome proliferator-activated receptors (PPAR) are transcription factors responsible for the transcription of genetic information important to the metabolic signaling network [11]. Three PPAR isoforms are known, i.e. α, γ and δ. PPARα is a regulator of fatty acid metabolism and energy homeostasis and is mainly expressed in the metabolically active tissues like liver, heart, skeletal muscle and adipose tissue. Molecules activating this receptor can reduce weight gain by increasing thermogenesis but also improve blood lipid levels important in several cardiovascular diseases [40]. PPARδ acts as central regulator of fat burning and thermogenesis. Activation of this isoform increases fatty acid oxidation and energy expenditure but has also shown to improve lipid and glucose levels in mouse models of type II diabetes [41]. PPARγ is mainly expressed in the adipose tissue and is a key mediator of adipogenesis. Activation of this receptor results in transforming fat storing white adipose tissue into fat burning brown adipose tissue and thus increased energy expenditure [42]. Furthermore, PPARγ activators were found to sensitize type II diabetes patients for insulin [43]. Taken together, agents that activate PPARs have great potential in obesity treatment by increasing energy expenditure and improving associated disorders.

1.4 Current Treatment Approaches for Obesity

Weight loss of just 5 to 10 % achieved by lifestyle intervention leads to significant improvements of cardiovascular disease risk factors [44]. However, most patients with obesity undergoing a lifestyle intervention will regain much of their lost weight on the long-term [45]. Another treatment option for patients with obesity is bariatric surgery. Although being one of the most effective treatments, bariatric surgery is reserved for patients with severe or complex obesity and additionally, carries risks of surgical complications and weight regain [46–49]. Combining lifestyle intervention with pharmacological treatment proved to be the most effective approach [50].

Currently, six drugs are approved by the federal drug administration (FDA) for treatment of obesity (Fig 3) [5]. Phentermine monotherapy is affecting the release of catecholamines in the hypothalamus causing reduced appetite and food consumption. Although phentermine has already been approved in 1959 for obesity management, the exact mechanism is still unclear [5]. The combination of phentermine with GABA-releasing properties of topiramate was found to be more effective than phentermine

monotherapy [51, 52]. Orlistat reduces systemic fat absorption by inhibition of gastric and pancreatic lipases, ultimately decreasing energy intake [32]. Lorcaserin is a selective agonist of the 5-HT$_{2C}$ receptor stimulating anorectic POMC/CART neurons [53]. The drugs naltrexone and bupropion act synergistically by antagonizing the µ opioid receptor and by inhibiting the reuptake of dopamine and norepiniphrine [54]. Liraglutide is the most recently approved obesity medication and acts via activation of the GLP-1 receptor increasing the insulin release from the pancreas [55].

According to clinical trials phentermine/topiramate is the most effective pharmaceutical obesity therapy causing on average an additional weight loss of only 7 % compared to placebo after one year of treatment [5]. Beside low efficacy, all currently approved medications show serious or at least unpleasant side effects underlining a dearth for efficient and safe anti-obesity medications [5].
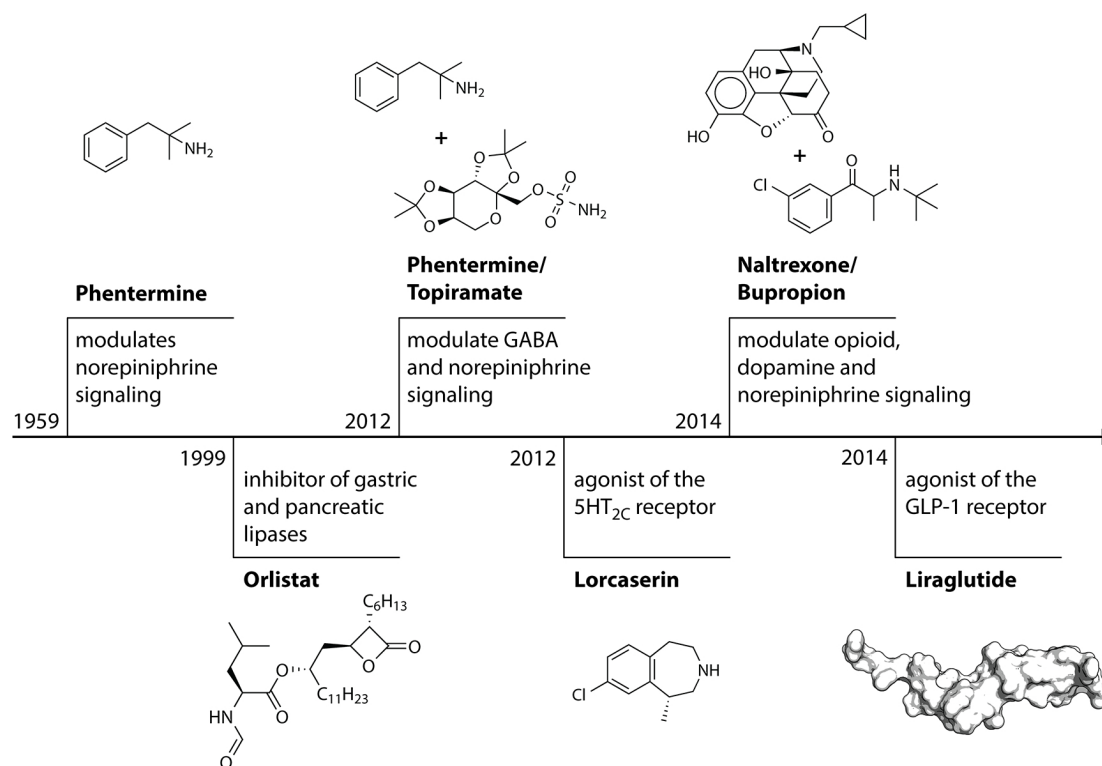
Figure 3: Medications currently approved by the federal drug administration for treatment of obesity [5].

1.5 GPCRs Targeted in This Thesis

Members of the G-protein coupled receptor (GPCR) family represent the key target for one third of FDA approved drugs rendering GPCRs an extremely important protein family for drug discovery [56]. Numerous hormones and neurotransmitters associated with obesity act via binding to GPCRs, e.g. ghrelin, NPY, AgRP and POMC [57].

Notably, except for orlistat, all currently approved anti-obesity medications target GPCRs.

The GPCR protein family consists of 800 members sharing a common motif of 7 transmembrane helices connected by alternating extracellular and intracellular loops (Fig 4A). They are responsible for transmitting extracellular signals into the cell and, ultimately, trigger cellular responses. The activity of GPCRs can be modulated by ligands with different size, ranging from photons, over small molecules like histamine to bigger peptides like NPY [57].

The binding of agonists induces a conformational change in the intracellular domain increasing the ligand-dependent recruitment of signal transducers like G-proteins, arrestins and extracellular signal-regulated kinases (ERKs) among others. In contrast, neutral antagonists block receptor signaling at the naturally imprinted basal level. Inverse agonists, can reduce the basal signaling level providing further therapeutic opportunities [58].

Fundamental research by Ballesteros and Weinstein on GPCR amino acid sequences has revealed conserved amino acid residues in each of the transmembrane helices [59]. This finding progressed into a generic residue numbering scheme that describes each residue of GPCR transmembrane helices with a helix number and a number indicating the distance to the most conserved residue of this helix [60]. By definition, the most conserved residue of a helix retrieves the number 50. All other residues of a helix are numbered according to the distance to this residue, e.g. the residue $V^{3.32}$ describes a valine of helix 3 that is 18 positions before the most conserved residue of helix 3 (Fig 4B). This numbering scheme has been proven extremely useful in comparing subtype specific effect of mutations and ligand interactions [60]. For instance, ligand binding to aminergic GPCRs commonly involves a charged interaction with residue $D^{3.32}$ [61]. Such information is indispensable for discovery of novel ligands targeting GPCRs.

1.5.1 Histamine H$_3$ Receptor

The histamine receptor family is a class of GPCRs with four known subtypes binding the endogenous biogenic amine histamine (Fig 5) [64]. The histamine H$_3$ receptor (H$_3$R) is mainly expressed in the CNS acting as presynaptic auto-receptor providing negative feedback for histamine release via G$_{i/o}$-protein dependent signaling. Additionally, H$_3$R was found to be involved in the regulation of other neurotransmitters including acetylcholine, norepinephrine, dopamine, glutamate, γ-aminobutyric acid and serotonin which raised interest for drug development against several neurologic and psychatric disorders [65]. The inverse agonist Pitolisant is the only currently approved drug that was developed to specifically target H$_3$R and is used for treatment of narcolepsy [66].

The histamine system was also investigated for its role in energy homeostasis and
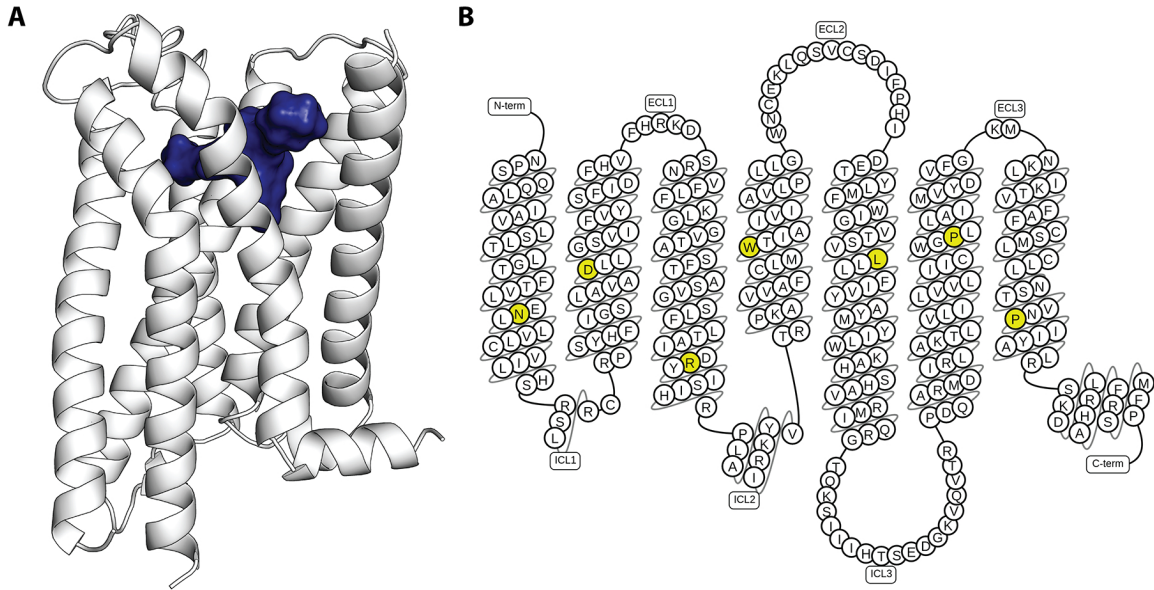
Figure 4: Structure of G-Protein coupled receptors. (A) 3D representation of cannabinoid receptor 1 crystal structure in complex with antagonist AM6538 (blue surface) [62]. (B) Snake plot of cannabinoid receptor 1 generated with a GPCRdb web service [63]. The most conserved residue for each transmembrane helix are highlighted in yellow.

as potential target of anti-obesity treatments. Administration of histamine reduces not only food intake but also increases thermogenesis in rodents [67]. In contrast, decreasing histamine levels induces body weight gain [68]. H_3R antagonists which block the negative feedback for histamine release showed anti-obesity effects in rodents, pigs and monkeys [69–73] (Fig 5). However, promising pre-clinical results could not be translated to human obesity treatment yet [74].

1.5.2   Melanin-Concentrating Hormone Receptor 1

The neurotransmitter melanin-concentrating hormone (MCH) and its G-protein coupled receptors MCHR1 and MCHR2 are expressed in several brain areas implicated in energy balance as well as sleep and arousal [75, 76]. MCHR1 signaling involves $G_{i/o}$ and G_q-proteins as well as ERK phosphorylation [77]. In contrast, MCHR2 was found to only recruit G_q-proteins [78].

Mice lacking either MCHR1, MCH or complete MCH neurons show increased loco-motor activity, have an increased energy expenditure and are resistant to diet-induced obesity [79–82]. The function of MCHR2 and its implication in energy balance is less well understood, since MCHR2 is not found in rodents complicating the study of animal models for obesity [83]. However, a transgenic mouse model indicates that MCHR2 might oppose the endogenous role of MCHR1, thus, favoring selective MCHR1 antagonists [84].

Several MCHR1 antagonists were already developed and showed promising pre-

clincial results in rodents [85–89] (Fig 5). However, non of the studied molecules was found to be effective in clinical trials in human [90]. Also, partially conflictive results have been published about the importance of MCHR1 antagonist CNS exposure. Several studies discontinued development of clinical candidates due to a low CNS exposure [89]. However, another study in rodents found an effective peptidomimetic antagonist selective for MCHR1 that is theoretically not able to enter the CNS [91]. The authors hypothesize that their studied MCHR1 antagonist is acting via binding to MCHR1 at adipocytes. These results underline the still limited understanding of pathways involved in obesity and demand further investigations.

H₃R ligands

Histamine   Pitolisant   A-331440

NNC 38-1049   GT-2394   Compound 17

MCHR1 ligands

MCH   GW803430   1j

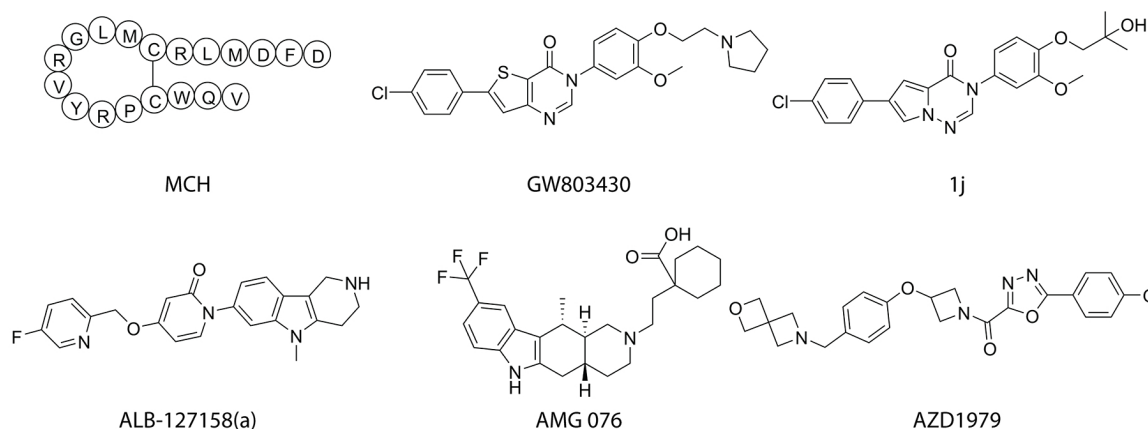ALB-127158(a)   AMG 076   AZD1979
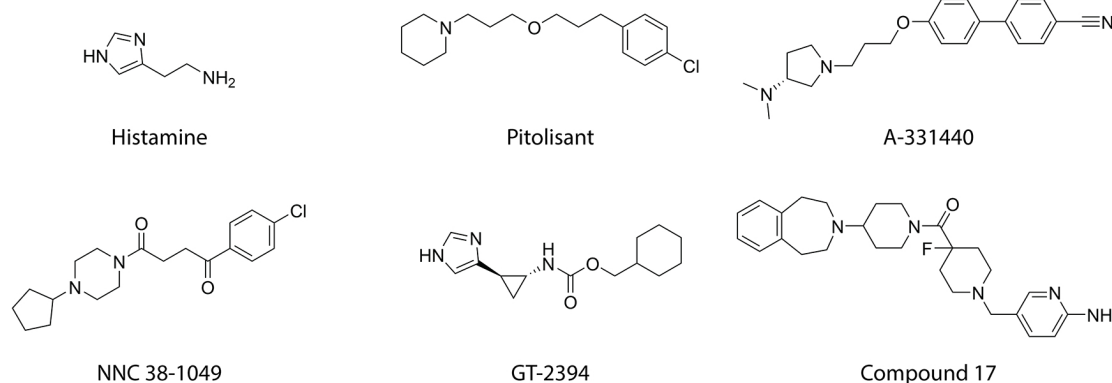
Figure 5: Endogenous ligands of H₃R and MCHR1 as well several drug candidates with promising pre-clinical results [67–73, 80, 85–89].

1.6   Rational Multi-Target Drug Design: An Emerging Paradigm

Driven by the experience that unwanted side effects often originate from binding to undesired off-targets, rational drug design traditionally focused on maximizing the

selectivity of ligands for a particular molecular target [92]. However, several "selective" drugs were later shown to be effective because of previously unknown activities towards other disease-relevant drug targets [93]. Also, our knowledge about possible target modulation by approved drugs is far from complete, since testing each drug against each known target is practically infeasible. Thus, many drugs that are being considered selective, may posses a yet unknown multi-target character [94].

Multi-target approaches question the dogma of maximizing selectivity for one target, suggesting that the effective adjustment of a phenotype might require the modulation of multiple targets, since biological networks can often find alternative routes to bypass the inhibition of a single target [92, 95]. Additionally, it is assumed that multi-target drugs can be used in smaller doses causing milder side effects, and furthermore decrease the possibility for drug resistance [95]. Especially drug design campaigns against complex diseases like obesity or psychiatric disorders that depend on several pathways may benefit from such approaches. Notably, most medications for mood disorders modulate the activity of multiple targets in the CNS [96].

Taken together, these characteristics render multi-target approaches a promising strategy to develop effective and safe anti-obesity pharmaceuticals.

Aim and Objectives

The constantly increasing prevalence of obesity indicates a large unmet of suitable medications to fight the obesity epidemic. Currently, available pharmaceuticals only show low efficacy and serious or at least unpleasant side effects. The complex etiology of obesity represents a challenge for traditional drug design campaigns and calls for novel approaches considering the multitude of pathways involved. This thesis aims at applying the multi-target concept to obesity drug development with the following steps:

- Identification of obesity-relevant target pairs that are modulated by molecules with high chemical similarity.

- Validation of the most promising target pair by *in vitro* evaluation of small molecules for potential multi-target character.

- Development of mechanistic structural 3D models to investigate key residues for ligand binding.

- Development as well as statistical and experimental validation of predictive 3D pharmacophores that allow for virtual screening of compound databases.

- Development of novel computational methods to overcome current limitations.

In this thesis, current methods in computational drug design were applied. The concepts behind these methods are described in the next chapter. The presented *in vitro* results for experimental validation were generated in close collaboration with the group of Prof. Dr. Holger Stark at the Heinrich-Heine-University in Düsseldorf.

Computational Methods

Computational techniques offer unique opportunities to accelerate the drug discovery process by providing access to highly efficient virtual screening as well as by rationalizing lead optimization efforts [97]. Depending on the employed data, computational methods can be classified into ligand- and structure-based approaches.

3.1 Ligand-Based Approaches

3.1.1 Molecular Similarity-Based Virtual Screening Methods

The utilization of molecular similarity measures in the drug discovery process is based on the observation that similar molecules often posses a similar property, e.g. affinity against a certain target [98]. Such measures can be categorized into three types, differing in the complexity of the underlying data, computational costs and accuracy (Tab 2).

The most simple, yet widely used similarity measure is the comparison of molecular properties. Lipinski's rule of five is a prominent example estimating the oral bioavailability of a molecule with the number of hydrogen bond donors and acceptors, molecular weight and computationally estimated octanol-water partition coefficient [99]. It has been intensively employed in the last two decades influencing decision making in the pharmaceutical industry. However, the concept of estimating oral bioavailability with molecular properties is nowadays being challenged [100].
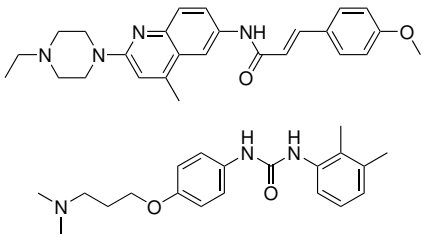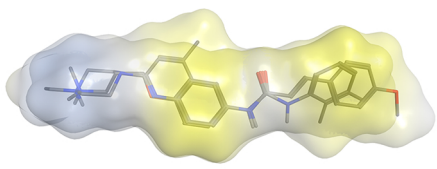
Initially being developed for substructure searches, molecular fingerprints formed the basis of additional methods for assessing the similarity of molecular structures on the two-dimensional level [101]. Most molecular fingerprints store the presence of chemical substructures in a binary vector. For example, the molecular access system (MACCS [102]) fingerprint searches for 166 substructures and sets corresponding bits to one if a substructure was found. In contrast, the extended-connectivity fingerprint records the identity and connectivity of each atom. This fingerprint was further developed classifying atoms into atom types (e.g. hydrogen bond donors and acceptors or hydrophobic and aromatic moieties) to increase the chance of scaffold hopping in similarity searches [103]. Corresponding bit vectors can be efficiently compared with

similarity coefficients, e.g. Tanimoto coefficient ranging from 0 (no similarity) to 1 (highest similarity) [104].

The development of molecular fingerprints enabled researchers to screen molecular databases for chemically similar molecules in a highly efficient manner [101]. Other prominent examples of fingerprint usage are the identification of protein targets binding similar molecules for multi-target or off-target analysis [105] and the identification of matched molecular pairs to rationalize structure-activity relationships [106].

Similarity measurements at the three-dimensional level are often less successful in retrospective virtual screening experiments than similarity measurements employing molecular fingerprints [107]. Also, 3D similarity is computational more expensive, since it requires the generation and comparison of multiple 3D conformations to achieve a reasonable performance. However, methods employing 3D similarity show more scaffold hopping, which is important to extend the chemical space of a hit series [107]. A prominent tool for assessing 3D similarity is ROCS (rapid overlay of chemical structures), which defines the volume of a molecule with atom-centered Gaussian functions [108]. The employed algorithms superpose conformations by maximizing the overlap of the molecule volumes. Additionally, implemented alignment and scoring routines analyze the chemical functionality of the molecules (e.g. hydrogen bond donors and acceptors or aromatic rings) to improve screening performance.

Table 2: Types of molecular similarity measures. HBA - hydrogen bond acceptor, HBD - hydrogen bond donor, MW - molecular weight, clogP - calculated octanol-water partition coefficient, Ro5 - rule of 5, ECFP - extended-connectivity fingerprint, ROCS - rapid overlay of chemical structures.

Type	Data				Tool
Molecular properties	HBD	HBA	MW	clogP	Ro5 [99]
	1	4	431	4.2	
	2	3	342	3.4	
2D structure					ECFP [103]
3D structure					ROCS [108]

3.2 Structure-Based Approaches

3.2.1 3D Models of Proteins at Atomistic Resolution

Atomistic models of proteins are essential for computer-aided drug design campaigns to understand the interactions formed between ligand and protein, and to suggest novel molecules for *in vitro* testing [97]. Advancements in nuclear magnetic resonance spectroscopy, X-ray crystallography and cryogenic electron microscopy drove the release of more than 140.000 experimentally resolved protein structures freely accessible at the Protein Data Bank [109, 110].

Despite this progress, many important drug targets still lack structural data. Especially membrane proteins such as GPCRs proved challenging to crystallize [111]. In such a situation, researchers rely on homology modeling, a method that exploits structural information from closely related proteins with resolved structures to computationally develop an atomistic model of the protein of interest [112]. Software tools like MODELLER [113] and MOE [114] provide algorithms to generate homology models with a defined workflow: (i) Template selection based on sequence similarity using efficient alignment algorithms like BLAST [115], (ii) alignment of template and target sequence with more accurate algorithms like Needleman-Wunsch [116], (iii) mutating residue mismatches into the corresponding residue of the target sequence, (iv) *de novo* modeling of target sequences without a corresponding template structure and (v) resolving severe atom clashes. Finally, specialized programs and web-servers like WHAT IF [117] can be employed to identify structural problems of the homology model, e.g. atom clashes or wrong dihedral angle distributions [Fig 6].
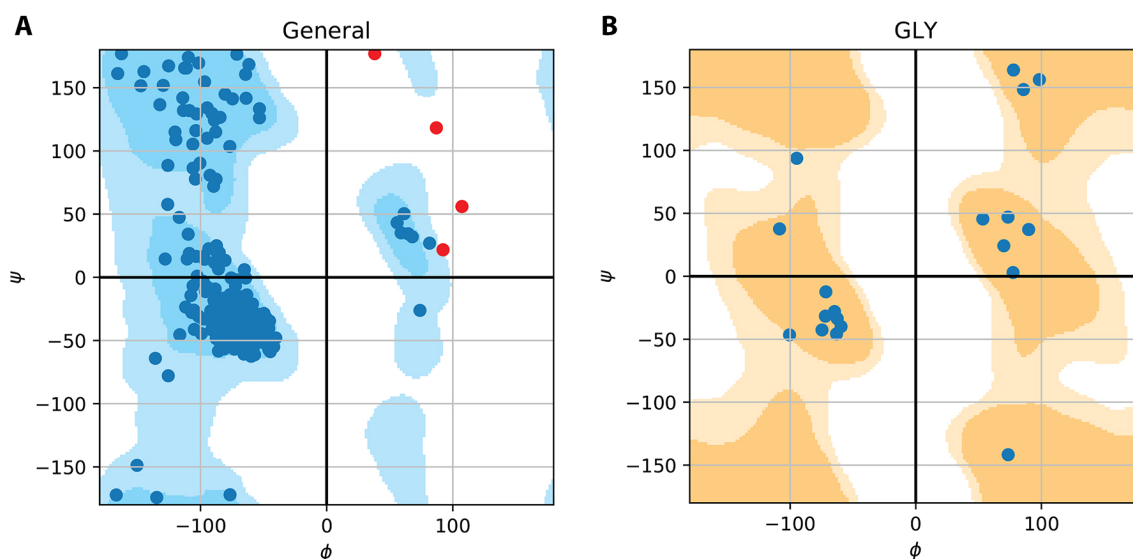

Figure 6: Ramachandran plots reporting the dihedral angle distribution for general (A) and glycine residues (B). Blue dots are inside the the area of dihedral angles observed in high resolution crystallographic structures, red dots outside.

3.2.2 Ligand-Protein Docking

Molecular docking is a frequently used method in structure-based drug design to predict the conformation of a small molecule inside a target binding site [118] (Fig 7). Several docking programs have been developed including GOLD [119], AutoDock [120], FlexX [121] and Glide [122]. Implemented algorithms sample the conformational space of ligands and evaluate the quality of each predicted pose with scoring functions [118]. GOLD, for instance, employs a genetic algorithm mimicking the process of natural selection by altering parameters based on mutation, crossover and migration [119]. Parameters describing the torsional, translational and rotational degrees of freedom are distributed over different chromosomes and each round of evolution results in a new combination of these parameters (docking pose), which is subsequently scored to allow a bias towards the fittest parameters. Scoring functions of docking programs significantly differ and typically involve the evaluation of several binding pose properties, e.g. steric and electrostatic complementarity, van der Waals attractive potential, desolvation energy and internal energy of the ligand [118].

Retrospective studies found that docking programs are successful in generating the binding conformation observed in experimentally resolved protein-ligand complexes [123, 124]. However, the calculated docking scores do not correlate with binding affinity for most of the studied targets. Despite this limitation, docking studies are an integral part of computational drug design campaigns aiding the rational explanation and exploitation of structure-activity relationships as well as the virtual screening of compound libraries for novel chemical entities [97].
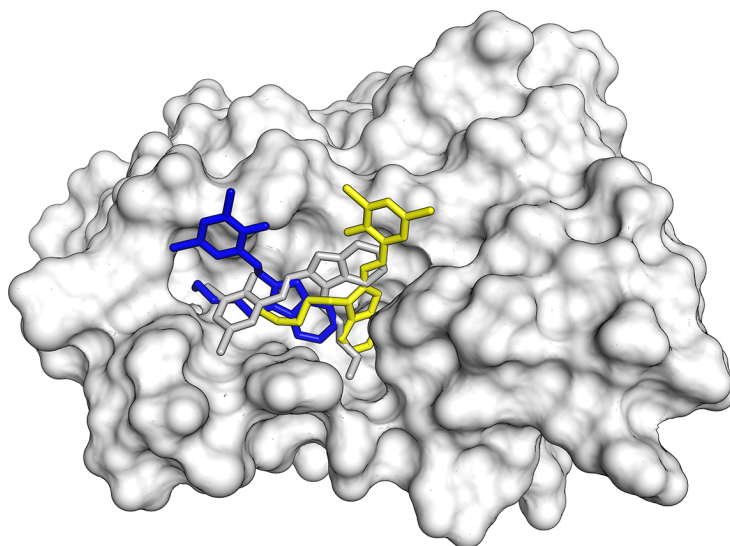
Figure 7: Predicted conformations of a small molecule inside a protein binding pocket.

3.2.3   3D Pharmacophores for Virtual Screening

The chemical space of small organic molecules for drug discovery is estimated to contain at least 10^{60} possible molecules [125]. The enormous improvements in computing power and artificial intelligence enables researchers to generate virtual compound libraries with billions of synthesizable molecules [126]. These resources can advance drug design campaigns but cannot be handled sufficiently with traditional *in vitro* high-throughput screenings [127]. Three-dimensional pharmacophore models describe the arrangement of electrostatic and chemical features required for a small molecule to bind its macromolecular target and can be used to screen such virtual compound libraries in a highly efficient manner [128]. Virtual screenings employing 3D pharmacophores typically achieve hit rates higher than 10 % rendering 3D pharmacophores an invaluable tool for drug design [129]. In the following review, the principles of 3D pharmacophores are described and recent developments in the field are highlighted.

Contribution:

Conceptual design (30 %)
Visualization (80 %)
Manuscript preparation (25 %)

ADVANCED REVIEW

WIREs
COMPUTATIONAL MOLECULAR SCIENCE   WILEY

Next generation 3D pharmacophore modeling

David Schaller | Dora Šribar | Theresa Noonan | Lihua Deng |
Trung Ngoc Nguyen | Szymon Pach | David Machalz |
Marcel Bermudez | Gerhard Wolber

Pharmaceutical and Medicinal Chemistry,
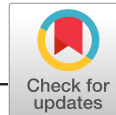Freie Universität Berlin, Berlin, Germany

Correspondence
Marcel Bermudez and Gerhard Wolber,
Pharmaceutical and Medicinal Chemistry,
Freie Universität Berlin, Königin-Luise-
Strasse 2+4, Berlin 14195, Germany.
Email: m.bermudez@fu-berlin.de;
gerhard.wolber@fu-berlin.de

Funding information
China Scholarship Council, Grant/Award
Number: 201606010328; Deutsche
Forschungsgemeinschaft, Grant/Award
Number: 407626949; Deutscher
Akademischer Austauschdienst;
Sonnenfeld Foundation

Abstract

3D pharmacophore models are three-dimensional ensembles of chemically
defined interactions of a ligand in its bioactive conformation. They represent an
elegant way to decipher chemically encoded ligand information and have there-
fore become a valuable tool in drug design. In this review, we provide an overview
on the basic concept of this method and summarize key studies for applying 3D
pharmacophore models in virtual screening and mechanistic studies for protein
functionality. Moreover, we discuss recent developments in the field. The combi-
nation of 3D pharmacophore models with molecular dynamics simulations could
be a quantum leap forward since these approaches consider macromolecule–
ligand interactions as dynamic and therefore show a physiologically relevant
interaction pattern. Other trends include the efficient usage of 3D pharmacophore
information in machine learning and artificial intelligence applications or freely
accessible web servers for 3D pharmacophore modeling. The recent developments
show that 3D pharmacophore modeling is a vibrant field with various applications
in drug discovery and beyond.

This article is categorized under:
Computer and Information Science > Chemoinformatics
Computer and Information Science > Computer Algorithms and
Programming
Molecular and Statistical Mechanics > Molecular Interactions

KEYWORDS

3D pharmacophores, artifical intelligence, machine learning, virtual screening, web services

1 | INTRODUCTION

Macromolecular biological structures such as proteins or DNA bind small organic molecules triggering functional modula-
tion and biological response. The way in which ligands bind to their macromolecular targets is based on a small set of chem-
ical interactions (chemical features), such as hydrogen bonds, charges, or lipophilic contacts. 3D pharmacophores represent

---

David Schaller, Dora Šribar, and Theresa Noonan contributed equally to this study.

an intuitive and powerful description of these interaction patterns. The high degree of abstraction in 3D pharmacophores enables the rationalization of binding modes for chemically diverse ligands and, subsequently, rapid and highly efficient virtual screening of molecular databases. Although the concept of 3D pharmacophores was developed at the beginning of the 19th century, virtual screening experiments were not performed until the late 80s and early 90s, when the first software packages for database searches were released.[1] The chemical space for molecules with a molecular weight below 500 Da is estimated to contain at least 10^{60} organic molecules.[2] Additionally, current developments in machine learning algorithms allow for in silico generation of billions of theoretically synthesizable molecules.[3] 3D pharmacophores present a unique opportunity to harvest the enormous available chemical space for drug-like molecules.

In this review, we give a comprehensive overview of 3D pharmacophore models, their usage in drug design, and current developments in the field. We introduce the basic concept and summarize the underlying methodology for describing binding modes and for applying 3D pharmacophore models in virtual screening. We highlight the power of 3D pharmacophore models in drug discovery by showcasing key studies for virtual screening as well as studies that aim at a mechanistic understanding of protein functions. Moreover, we present and discuss current developments such as the integration of molecular dynamics, the combination with machine learning, and freely accessible web services.

2 | THE PRINCIPLES OF 3D PHARMACOPHORES

3D pharmacophores capture the nature and three-dimensional arrangement of chemical functionalities in ligands that are relevant for molecular interactions with the macromolecular target. Chemical functionalities are thereby classified into more general pharmacophore features, for example, hydrophobic areas, aromatic ring systems, hydrogen bond acceptors, hydrogen bond donors, negatively ionizable groups, and positively ionizable groups.[4] Less common interaction types that contribute to the binding of ligands, such as metal coordination and halogen bonds, are either already implemented in most software packages or require user definition.[5–7] Besides chemical nature and spatial arrangement, 3D pharmacophores can capture feature directionality in the case of hydrogen bonds and aromatic interactions.[8] Additionally, spatial tolerance and weight can be fine-tuned for each pharmacophore feature to adjust its size and importance in the 3D pharmacophore. In order to describe the preferable shape of molecules in the binding site, pharmacophore features are often combined with exclusion volume constraints (also referred to as excluded volume constraints). For instance, an exclusion volume constraint may consist of a set of spheres that represent the protein residues imposing a barrier for binding of potential ligands.

Several 3D pharmacophore modeling programs have been developed, of which several are free for academic users (Table 1). Although the exact definition and implementation of pharmacophore features and their characteristics may differ between different 3D pharmacophore modeling programs, the underlying concept of 3D pharmacophores remains the same.

2.1 | 3D pharmacophore elucidation

3D pharmacophore elucidation methods can be classified as feature-based, substructure pattern-based, or molecular field-based, depending on how the pharmacophore features are derived. Feature-based methods derive pharmacophore features by filtering for geometric descriptors that match the characteristics of molecular interactions. Pattern-based methods, such as those implemented in PHASE, LigandScout, and Catalyst, detect substructures for chemical features in molecules. For example, all hydroxyl groups are defined as hydrogen bond donors and acceptors. In contrast, molecular field-based methods such as FLAP and Forge sample the molecular surface of either ligand or macromolecular target with different chemical probes and calculate interaction energy maps which can be translated into pharmacophore features. An additional distinction between 3D pharmacophore generation methods is based on the type of employed data. This could be a set of active ligands, structural data on the ligand in complex with its macromolecular target, or structural data of the macromolecular target alone (Figure 1).

2.1.1 | Ligand-based 3D pharmacophores

Ligand-based 3D pharmacophores are used when no structural information on the macromolecular target is available. They are composed of chemical features shared by a set of active compounds that are important for the interaction with

TABLE 1 3D pharmacophore modeling software, their components, and availability of free academic licenses

Software	Input	Identification methods	Virtual screening capability	Free for academic use[a]
FLAP[9]	Ligand, complex, apo	Molecular field	Yes	No
Pharmer[10]	Ligand, complex	Substructure pattern, feature	Yes	Yes (GPLv2)
LigandScout[11]	Ligand, complex, apo	Substructure pattern, feature, molecular field	Yes	No
Catalyst[12]	Ligand, complex, apo	Substructure pattern, feature, molecular field	Yes	No
MOE[13]	Ligand, complex, apo	Substructure pattern, feature, molecular field	Yes	No
PHASE[14]	Ligand, complex, apo	Substructure pattern, feature, molecular field	Yes	No
Pharao[15]	Ligand	Substructure pattern	Yes	Yes (GPLv2)
UNITY[16]	Ligand, complex	Substructure pattern, feature	Yes	No
Forge[17]	Ligand	Molecular field	Yes	Free for PhD students

Note: Software name describes virtual screening component. Names of 3D pharmacophore generation components in respective software suite may vary. 3D pharmacophores are commonly derived from an overlay of small molecules (ligand), from a ligand bound to its macromolecular target (complex) or from the macromolecular target alone (apo). Pharmacophore features are identified by analyzing molecular fields describing the potential interaction energy toward molecular probes (molecular field), by searching for substructure patterns able to perform the respective interaction (substructure pattern) or by filtering for geometric descriptors fulfilling the criteria for molecular interactions (feature).

[a]The software website was searched, or holder of rights was asked for academic license conditions.

FIGURE 1 3D pharmacophore generation approaches based on the available data. 3D pharmacophores can be generated from either a set of known ligands, atomistic models of ligand-macromolecular target complexes or the sole macromolecular (apo) target

Ligand   Complex   Apo
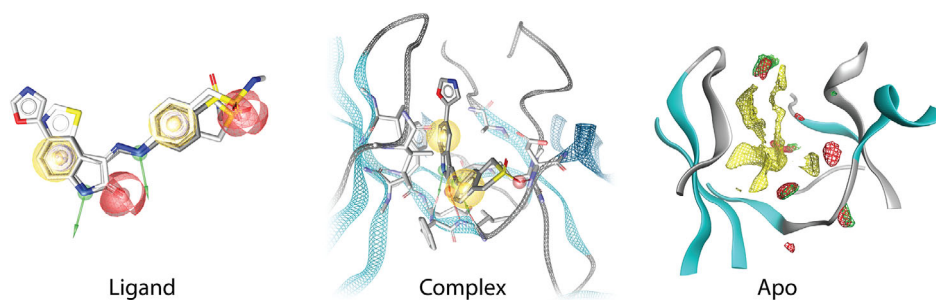
the target (Figure 1). Shared pharmacophore features are usually derived from the 3D alignment of different conformations of active compounds. 3D structures of the conformers are aligned so that the same pharmacophoric features are located in similar positions. If all the aligned molecules share a certain feature at a specific position, a pharmacophoric feature is placed at this position.[18] 3D alignments are often preceded by prefiltering steps based on quick distance checks, which substantially reduce computational time. For instance, the HipHop algorithm in Catalyst uses "pruned exhaustive search" and gradually builds-up shared 3D pharmacophores from the two-feature pharmacophores found in conformers of molecules.[12] In order to identify a shared 3D pharmacophore at each step, a precomputed list of all the interfeature distances in the molecule is first checked to see whether the specific feature combination is present. This prefiltering step is followed by alignment by least-squares fit of the features. LigandScout identifies optimal alignment by first checking best pairings between two sets of pharmacophore features based on interfeature distances, followed by alignment using the Kabsch algorithm.[19] In some software packages, such as HypoGen in Catalyst, the derived three-dimensional arrangement of chemical features can be correlated with biological activities of known actives.[14,20,21] This step can help to determine the importance of each feature for small molecule bioactivity.

However, it is important to note that bioactive conformations of the molecules are usually not known. Therefore, ligand-based 3D pharmacophore software considers a set of low energy conformations for each molecule. Although commercial conformer generation algorithms are generally successful in reproducing bioactive conformations, the ligand-based 3D pharmacophore generation procedure is not guaranteed to yield an alignment with the bioactive conformations.[22] Another limitation of ligand-based 3D pharmacophores is the dependence on structurally similar molecules, since structurally more diverse molecules may not share the same binding mode and hence, require separate pharmacophore models. But even if different molecules share a common binding mode, a correct alignment becomes more challenging to the ligand-based algorithms the more diverse the molecules are.[8]

2.1.2 | Structure-based 3D pharmacophores

Structure-based 3D pharmacophore elucidation can be performed on atomistic models of two types of structures. In macromolecule–ligand complexes, a ligand is present in the binding site of the target molecule (Figure 1). Ligands in complex with a macromolecular structure are primarily either co-crystallized or docked into the target site. If there is no available structure of a macromolecule–ligand complex, or no known ligands at all for a binding site, programs can derive 3D pharmacophore models from atomistic models of apo structures (Figure 1; Table 1). These apo structures are atomistic models of macromolecules bound to no ligand.

Apo 3D pharmacophore elucidation techniques are especially useful in cases where there are no known ligands, necessitating a de novo approach to pharmacophore feature placement within the cavity. However, apo 3D pharmacophore generation methods can also be useful when applied to structures of macromolecule–ligand complexes. In this instance, they can be used to generate a novel 3D pharmacophore for the same active site that is unbiased by the existing ligand. This can be used to explore a novel region of chemical space for the same binding cavity. Accordingly, one of the strengths of 3D pharmacophore-based virtual screening is the potential for scaffold hopping afforded by the arrangement of abstract features not bound to any specific ligand structure.

Though they are the most common drug targets, proteins are not the only macromolecular structures analyzed in 3D pharmacophore development. Programs including LigandScout and Catalyst (Table 1) can generate 3D pharmacophore models based on nucleic acids. For example, Spitzer et al. generated a 3D pharmacophore hypothesis for minor groove binders based on a DNA–ligand complex.[23]

Feature-based methods can be employed on macromolecule–ligand complexes as well as on empty binding sites. Feature-based programs analyze a target–ligand complex and employ a set of chemical and geometric rules to identify and classify target–ligand interactions, which then comprise the pharmacophore features.[19] In an example of a feature-based method being applied to an apo structure, a strategy was developed by Schrödinger whereby fragments are docked into an apo binding site using the Glide XP docking program.[24,25] The most energetically favorable fragment docking poses are selected to construct the 3D pharmacophore hypothesis using Phase (Table 1).[14]

Molecular field-based methods, such as FLAP (Table 1), employ molecular interaction fields (MIFs) to identify hotspots for pharmacophore feature placement.[9] A prominent tool for generating MIFs is the GRID software, which is well known for its role in the discovery of the antiviral drug zanamivir.[26,27] In principle, an evenly spaced grid is placed over a predefined binding cavity, and probes are placed to sample the binding site. These probes take the form of moieties representing the interactions most likely to occur between the macromolecule and ligand functional groups. As a next step, the energy between probe and target structure atom is calculated to define interaction sites. Thus, these probes can identify sites of favorable interactions with the macromolecule. These interaction energies generate MIFs, which are contoured by energy to generate maps that describe how the interaction energy between the target and a given probe varies over the surface of the target. Molecular field-based programs take the points where the energy of a MIF represents a local minimum, termed "hotspots," and convert them into pharmacophore features according to the type of probe that forms the most energetically favorable interaction at this point. Molecular field-based hotspot detection can also be performed by employing noncommercial software such as AutoGrid within AutoDock which provides access to energy grid maps for various atom types.[28,29]

Programs creating pharmacophore features for apo binding sites generate a surplus of possible features. These must be reduced to a reasonable number for virtual screening; a balance between enough features to allow for specificity, but not too many features, as this would be too restrictive and could lead to false negatives. Some programs include a feature reduction functionality, but other programs output an initial, unrefined 3D pharmacophore. The initial unrefined 3D pharmacophore composed of many features must then be reduced by the user. Feature selection can be based on information about the binding site and binding site-lining atoms, and according to which features of the binding site the user would like the ligands to exploit. Feature reduction must not only be performed manually; HS-Pharm[30] is an example of a program that uses machine learning to reduce the number of initial 3D pharmacophore features, as discussed in the advanced section later.

2.2 | Pharmacophore-based virtual screening

In pharmacophore-based virtual screening, 3D pharmacophores developed from either a set of active ligands, a target–ligand complex or the apo target, are screened against virtual libraries of molecules. Molecules that satisfy the query

pharmacophore requirements are retrieved from the libraries. The prioritization of compounds by virtual screening can dramatically increase the hit rate compared to in vitro high-throughput screenings and hence reduce the number of compounds for experimental testing (Figure 2).

To address conformational flexibility of the molecules, conformer libraries for the screened compounds are prepared before the screening step. It is worth mentioning that conformation generation is handled differently by screening software packages. Some software packages, such as LigandScout, Catalyst, or MOE perform virtual screening on a pre-generated set of conformations for each library molecule, while other software packages, such as PHASE, allow on-the-fly conformer generation during the screening step sacrificing virtual screening speed.[19,31,32] For more information on conformer generation for virtual screening, the reader is encouraged to read available publications on this topic.[8,22,32–36]

In the screening step, pharmacophoric features in the query pharmacophore are compared to pharmacophoric features present in the molecules of the screened library. Comparison methods can be divided into two distinct approaches: fingerprint-based and 3D alignment-based. Fingerprint-based methods, such as FLAP, primarily extract information about feature presence and/or interfeature geometries into fingerprint-like descriptors, which enables time-efficient similarity (e.g., using the Tanimoto coefficient) comparison between the query pharmacophore and the conformer library. Alignment-based methods including LigandScout, Catalyst, and PHASE perform 3D alignment of the pharmacophore feature set. A match is reported if the pharmacophoric feature set of a distinct conformation of a molecule can be aligned to the feature set of the query pharmacophore. 3D alignment is computationally expensive and time consuming, especially in the context of large molecular library screening. In order to reduce computational time, 3D alignment is often preceded by a fast prefiltering step based on feature-types, feature-counts, or fast distance checks.
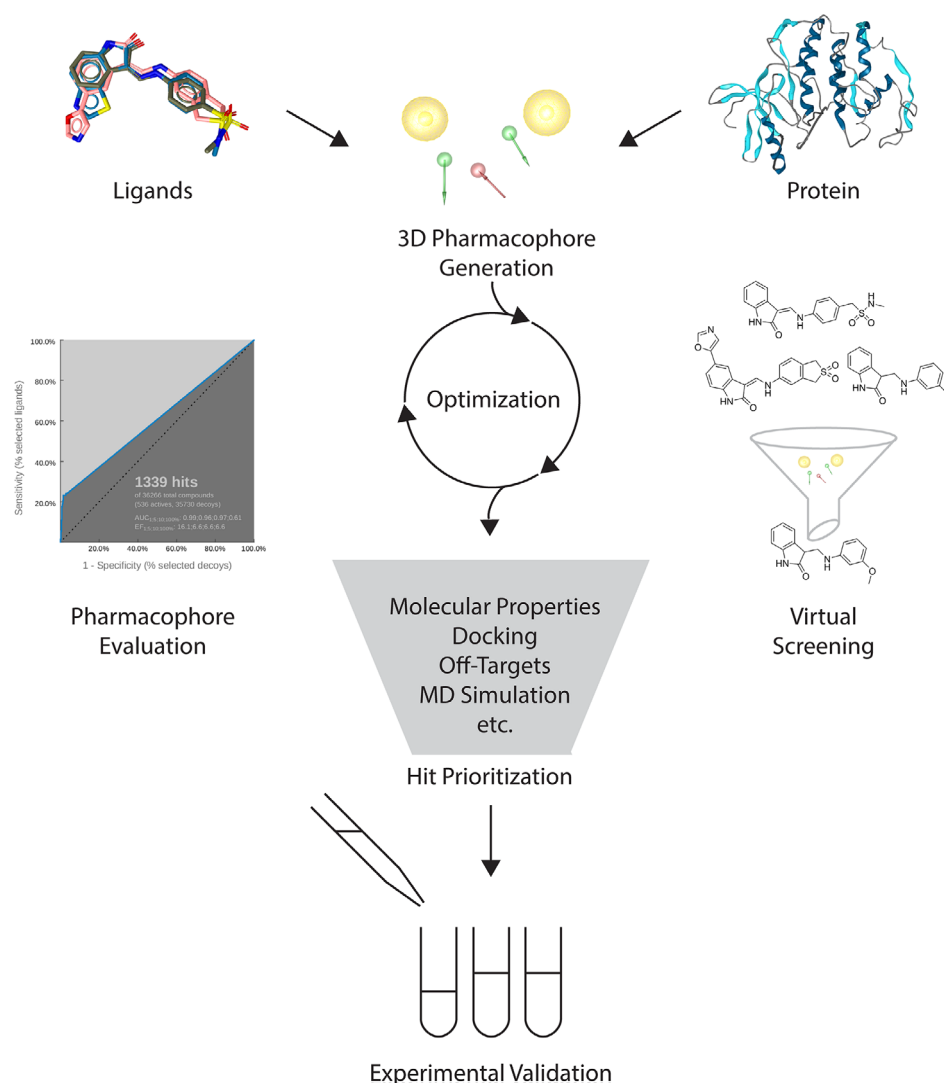
FIGURE 2 Virtual screening workflow. 3D pharmacophores are generated with either structure- or ligand-based approaches. State-of-the art retrospective validation is performed by plotting ROC curves with elaborated sets of actives and decoys. Pharmacophore-based virtual screening is often followed by computationally more expensive methods such as docking or molecular dynamics simulations to get more differentiated structural insights. ROC, receiver operating characteristics

LigandScout is the only program that provides loss-less prefiltering steps, providing the most geometrically most accurate screening algorithm. Additionally, its unique pattern-matching 3D alignment algorithm results in screening hit lists that are orthogonal to other programs, mainly relying on interfeature distance fingerprints.[37]

In cases where experimental data for binding ligands is available, derived 3D pharmacophore models can be validated. Usually, a validation set for 3D pharmacophores contains reported active, inactive and decoy molecules. Two points should be taken into consideration when preparing the validation set: Firstly, 3D pharmacophores describe a unique binding pose. Therefore, the active set should include molecules that share the same binding mode within the target protein. Secondly, reported inactives should be included with caution, as observed inactivity may result from other factors, for example, insolubility or inability to reach the target in cell-based assays. Therefore, the use of carefully selected decoys is encouraged over inactive molecules. A decoy is a compound presumed to be inactive and showing a high similarity in physicochemical properties to the active compounds. The Directory of Useful Decoys (DUD-E) provides a convenient web-based tool for the generation of decoys.[38] Subsequent screening against a validation set can be used to assess the quality of the developed 3D pharmacophore, and to further optimize it (Figure 2).

When evaluating the quality of a 3D pharmacophore model by its performance in virtual screening, various metrics, or enrichment parameters, are employed (Figure 2). 3D pharmacophore performance is evaluated in terms of how many actives the 3D pharmacophore can retrieve from a data set, and how well the 3D pharmacophore is able to correctly classify compounds as active or inactive. Enrichment parameters classify the compounds in the data set into one of four categories: active (true positive, TP); inactive but identified as active (false positive, FP); inactive (true negative, TN); active but classed as inactive (false negative, FN). Different metrics measure different aspects of 3D pharmacophore performance. These metrics include the yield of actives (YA), which describes the number of true positives present in the list of total hits retrieved by the 3D pharmacophore.[39] Receiver operating characteristics (ROC) curves plot the rate of true positives identified over the rate of false positives, thus displaying the sensitivity and specificity of the 3D pharmacophore model, characterizing how many active hits the 3D pharmacophore can identify in relation to how many inactive compounds it misidentifies as active.[39–41] For a comprehensive list of enrichment parameters, how they are calculated, and their uses, the reader is referred to Braga and Andrade.[40]

After a 3D pharmacophore has been developed and retrospectively validated, it can be used to screen available commercial or in-house compound libraries (Figure 2). Depending on the complexity of the 3D pharmacophore and the size of the library, hit lists of various sizes will be retrieved by pharmacophore screening. 3D pharmacophore-based screening is often followed by further characterization of the binding mode with methods like molecular docking, targeted molecular dynamics simulations, or other methods to gain more structural information to rationally prioritize molecules for experimental testing (Figure 2).

3 | APPLICATION CASE STUDIES

3.1 | Virtual screening

Besides molecular docking, 3D pharmacophores are widely applied for virtual screening. In this section, we highlight and discuss recent success stories of virtual screening campaigns covering different target classes and methodologies.

3.1.1 | Balancing the immune system with small molecule modulators

Toll-like receptors (TLRs) act as key players in the activation of the innate immune response by recognizing molecular patterns associated with infections and nonphysiological tissue damage.[42] Rational design of small molecule TLR modulators is a promising strategy to treat autoimmune inflammation, cancer, or allergies, or to identify adjuvants for vaccines.[43]

In 2014, Murgueitio and colleagues were facing a sparse data scenario with no small organic TLR2 inhibitors available.[44] Therefore, they generated a 3D pharmacophore based on MIFs to define key interactions necessary for ligand binding. A structure-based pharmacophore model was carefully developed. Subsequent virtual screening revealed novel antagonists in the low micromolar range with biological activity for 20% of their virtual hits. With more small organic TLR2 ligands reported later on, Murgueitio and colleagues continued their efforts in searching for novel TLR2

modulators and used a combined 3D pharmacophore and shape-based approach to discover a novel pyrogallol-based compound (MMG-11) through virtual screening.[45,46]

In 2019, Šribar and co-workers conducted a structure-based virtual screening followed by in vitro experimental validation in seeking novel TLR8 modulators.[47] Molecular docking was performed to explore different binding modes able to explain the activity of known modulators. The most descriptive binding mode was translated into a 3D pharmacophore that was subsequently employed for virtual screening. This approach finally led to a novel chemotype for TLR8 inhibitors, where 36% of the molecules retrieved by this virtual screening approach showed activity in vitro.

3.1.2 | Discovery of novel covalent binding ligands

The design of covalent-binding ligands has gained increased popularity in the drug discovery community.[48,49] Prolonged residence time on targets and pharmacodynamic effects independent from pharmacokinetics allow for lower doses or longer dosage intervals, making covalently binding drugs attractive for many therapies.[50,51] Covalent docking and quantum mechanics (QM) calculations represent the "gold-standard" for the development of covalent binders.[52,53] Docking is suitable for screening of small compound libraries, whereas QM calculations can be applied for single compounds only, due to high computational costs and time. For virtual screening of large databases, a pharmacophore-based approach is more applicable. Schulz and colleagues introduced a novel feature called "residue bonding point," which recognizes drug-like warheads, such as ketones, nitriles, or Michael acceptors, to the LigandScout framework.[11,54]

They employed the residue bonding point feature, also referred to as "covalent feature," for the de novo design of viral 3C protease inhibitors.[54] A selective and specific 3D pharmacophore was generated that included noncovalent interactions crucial for substrate-recognition, as well as the novel covalent feature. The obtained 3D pharmacophore was used for virtual screening of a fragment library. Compounds showing a high similarity of docking poses in covalent and noncovalent form were selected for in vitro testing. Compound **F1**, a heterocyclic aromatic ketone, showed the highest inhibitory activity in an enzymatic assay (Figure 3). The covalent binding to the Coxsackievirus (CV) B3 protease was proven with protein mass spectrometry. Compound **F1** was optimized using a scaffold-hopping strategy, yielding in the more stable and active hit **C5**, a phenylthiomethyl ketone. This compound was modified using synthetic approaches to produce **7a**, a selective and irreversible covalent inhibitor of CV B3 and Enterovirus (EV) D68 protease. This example illustrates the ability of 3D pharmacophore models to not only identify novel ligands, but also their suitability for hit and lead optimization. Additionally, this study highlights the applicability of 3D pharmacophores for increasingly popular fragment-based drug discovery campaigns.[55]

3.1.3 | Targeting GPCRs with 3D pharmacophores

G protein-coupled receptors (GPCR) are important drug targets due to their omnipresence in human tissues, accessibility to drugs, and regulatory roles in many physiological and pathophysiological processes.[56] Therefore, they are widely targeted by virtual screening campaigns in search of novel bioactive ligands.

In 2017, Frandsen and colleagues used an in-house developed method to build a histamine H3 receptor (H3R) pharmacophore model for virtual screening campaigns.[57] Ligand-residue fragments were extracted from available GPCR crystal structures and mapped to the same conserved binding pocket residues of the target receptor.[58] This method allows for structure-based modeling of orphan receptors with insufficient structural data or unknown ligands. Due to its reliance on existing ligand-receptor fragments from only 62 GPCR crystal structures, the initial H3R pharmacophore model missed an important cationic feature, which was added through docking studies and the matching of known H3R ligands to the apo pharmacophore model. Pharmacophore features were placed with Phase.[14] Virtual screening, hit selection, and generation of analogue from potent ligands amounted to 76 compounds being pharmacologically tested with an IP1 accumulation and radioligand binding assay. Five neutral antagonists and one inverse agonist showed binding in the low micromolar range, resulting in a hit rate of 8%.

Another approach was followed by Schaller and co-workers using a ligand-guided homology modeling strategy to discover novel H3R ligands.[59] A key aspect of this approach is the prioritization of a set of 1,000 homology models based on their ability to explain the binding of nine known antagonists. The selected homology model was subsequently used for 3D pharmacophore model generation with LigandScout.[11] Complementarily, 10 diverse H3 receptor ligands were
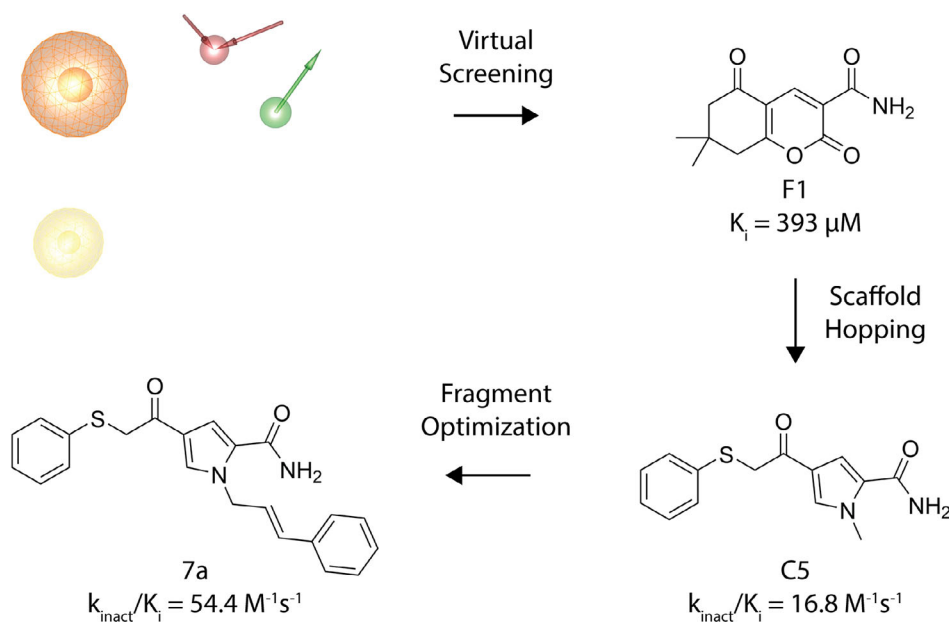
FIGURE 3 Discovery of covalent inhibitors of viral 3C protease. The initial fragment was identified with a 3D pharmacophore and further optimized by scaffold hopping and subsequent fragment growing. Green arrow—hydrogen bond donor, red arrow—hydrogen bond acceptor, yellow sphere—lipophilic contact, orange sphere—residue binding point

docked and evaluated by their interaction pattern. Resulting 3D pharmacophore models were iteratively optimized and validated with a set of 100 diverse active compounds and 3,051 decoys, using the three best performing 3D pharmacophore models for parallel virtual screening. Subsequently, eight hit molecules were selected for biological testing of which two showed nanomolar affinities in a radioligand depletion assay, resulting in a hit rate of 25%.

3.1.4 | Design of multitarget ligands for controlling inflammation

The arachidonic acid (AA) cascade is a key biochemical pathway for the inflammatory response involving the production of pro-inflammatory lipid mediators such as leukotrienes via 5-lipoxygenase-activating (FLA) protein, but also for anti-inflammatory mediators like epoxyeicosatrienoic acids. The latter are hydrolyzed by soluble epoxide hydrolase (sEH). The simultaneous inhibition of both enzymes therefore represents a promising approach for controlling inflammation mediators derived from AA. Schuster and co-workers applied a pharmacophore-based virtual screening to discover the first dual inhibitor of FLA protein and sHE with activities in the nanomolar range.[60] In a first step, they virtually screened the SPECS library with two different ligand-based pharmacophore models both derived from known FLA protein inhibitors. Twenty selected hit molecules were further prioritized by previously reported structure-based sHE models and resulted in one novel and potent dual FLA protein/sHE inhibitor.[61] Since multitarget approaches are getting more and more attention, this example shows that ligand-based pharmacophores could have some benefits when applying on multiple targets that bind chemically similar physiological ligands. Moreover, it highlights the potential of combining ligand and structure-based models in multitarget approaches, in which one 3D pharmacophore serves as a prioritization tool for the hitlist derived by the other model.

3.1.5 | Design and optimization of novel agents in crop science

A successful application of ligand-based design in crop sciences was demonstrated by Yao et al. by targeting an enzyme in plant-pathogenic fungi.[62] The group performed pharmacophore-based virtual screening to find inhibitors of succinate dehydrogenase (SDH). This protein is involved in the electron transport chain in eukaryotic mitochondria and represents a validated target for pesticides. All marketed SDH inhibitors consist of a carboxylate and amine-moiety coupled to an aromatic amide. Yao et al. generated a ligand-based 3D pharmacophore containing aromatic, lipophilic, and amide features applying Catalyst. The 3D pharmacophore was validated and used for a virtual screening campaign of a focused amide library developed in-house. To construct the focused amide library, diverse commercially available carboxylate moieties were linked to aniline using Discovery Studio's *Enumerate Library by Reaction* protocol to explore

the chemical space of carboxylate-cores. After virtual screening of the focused library, 16 compounds were selected for in vivo testing. Eight compounds showed more than 50% inhibition when tested against three different fungi species at a concentration of 100 mg/L. The ligand showing the highest and broadest activity was optimized via synthesis. In a second focused amide library, diverse amine moieties were linked to the best carboxylate core and screened against the 3D pharmacophore. The resulting collection of derivatives was tested in vitro on SDH. By combining in silico and experimental optimization steps, a broadly active novel amidic SDH inhibitor with low micromolar activity was developed.

3.2 | Understanding protein functionality

Besides describing ligand binding modes and virtual screening applications, 3D pharmacophores are powerful tools to investigate ligand-dependent protein functionality on a mechanistic level. The following paragraphs showcase examples where 3D pharmacophore models play an essential role in contributing to the mechanistic understanding of pharmacological effects.

3.2.1 | Modeling metabolism

Sulfotransferases (SULTs) play an important role in phase II metabolism, but represent challenging targets due to their high flexibility and broad substrate specificity. Rakers and colleagues developed a pharmacophore-based SULT prediction model to discriminate between substrates, inhibitors, and ligands that show both characteristics dependent on their concentration.[63] This study is remarkable for two reasons. Firstly, it uses an ensemble of different enzyme conformations derived from molecular dynamics (MD) simulations to generate conformation-specific 3D pharmacophore models. Secondly, the pharmacophore fit score was incorporated into a machine learning approach based on support-vector machines for post-filtering of screening results. The resulting pharmacophore-based prediction model was successfully applied to the screening and classification of ligand types for SULT1E1 and enhances our understanding of SULT enzyme specificity.

The aryl hydrocarbon receptor (AHR) is a ligand-dependent transcription factor controlling the metabolism of physiological substances and xenobiotics. Based on carefully validated homology models, Tkachenko and colleagues applied 3D pharmacophore models to study differences between physiological ligands and xenobiotics with regard to AHR transport to the nucleus and subsequent induction of CYP1A1.[64] Histidine 291 was identified as a key residue, which controls both functionalities, but with different roles in binding of physiological ligands such as kynurenine and xenobiotics such as β-naphtoflavone.

3.2.2 | Investigating ligand-dependent receptor function

GPCRs represent an important drug target class with highly complex pharmacology and various possibilities to modulate receptor function in a ligand-dependent manner. One major issue in this field is receptor selectivity, especially for closely related subtypes of the same family. In order to understand subtype selectivity of bitopic (dualsteric) ligands at muscarinic receptors, Bermudez and co-workers used 3D pharmacophore models to identify subtype-specific interaction patterns.[65] This rationally explained, on the one hand, how ligands achieve selectivity for a certain subtype, and, on the other, identified key residues in the extracellular loop regions (e.g., a M3-specific salt bridge) that account for subtype-specific receptor functionality. Some of the aforementioned bitopic ligands for the M2 receptor showed some unexpected yet interesting pharmacological properties, such as partial agonism and pathway-specific receptor activation (biased signaling). In order to understand these effects, 3D pharmacophore models were combined with other modeling techniques and pharmacological experiments.[66,67] The partial agonism could be explained by the existence of multiple binding modes, which stabilize different activation states of the receptor. This concept was validated by experiments and resulted in the rational design of a full agonist, which can only adopt the binding mode stabilizing active receptor states.[66] The pathway-specificity of the biased ligands was studied in a similar setting and resulted in a mechanistic model whose key concept resides in the conformational restriction of the extracellular loop region.[67–69] In another study, the effect of fluorination of the photoswitchable azobenzene core was investigated in muscarinic agonists. This

study shows that fluorination of the photoswitch alters not only the photochromic behavior but also the pharmacological profile at the M1 receptor, due to additional interaction possibilities.[70] In all of these GPCR studies, 3D pharmacophores were key to understanding ligand-dependent receptor functions and, moreover, turned out to serve as the optimal instrument for communication with synthetic chemists and pharmacologists.

4 | ADVANCED APPROACHES EMPLOYING 3D PHARMACOPHORE PRINCIPLES

In the previous sections, we gave an overview of the concept of 3D pharmacophores and on which established software is available. Additionally, we presented several state-of-the-art application case studies employing 3D pharmacophores for prospective virtual screening and for understanding protein functionality. But what could be considered an advanced approach? Typically, 3D pharmacophores are generated from atomistic models of macromolecules or from an alignment of multiple ligand conformations and are employed for analysis of structure–activity relationships or virtual screening campaigns. Moreover, a local installation of software and the availability of a high-performance computer are usually mandatory to perform virtual screening experiments. In the following section, we introduce advanced approaches that integrate conformations from MD simulations, employ machine learning algorithms, and provide access to 3D pharmacophore searches without the requirement of expensive licenses and high-performance computers (Table 2).

4.1 | Integration of information from molecular dynamics simulations

Since both macromolecules and ligands are dynamic entities, it becomes apparent that this also holds true for macromolecule–ligand complexes and the underlying interactions. Following this idea, Carlson and colleagues integrated information from MD simulations in the development of an enhanced 3D pharmacophore model to virtually screen for novel HIV-1 integrase inhibitors.[92] Later, Carlson used the HIV-1 protease to show that a 3D pharmacophore generated from an ensemble of 28 NMR conformations performs better than a 3D pharmacophore generated from 90 X-ray structures.[93] This pioneering work inspired other researchers and kick-started the development of several methods employing conformations from MD simulations for 3D pharmacophore generation.

Hydration-site-restricted pharmacophore (*2012*). Unrefined 3D pharmacophore models generated from apo binding cavities usually contain too many features for efficient virtual screening. The hydration-site-restricted pharmacophore (HSRP) approach aims at reducing the number of pharmacophore features by identifying hydration sites on the protein surface, whose water molecules suffer from unfavorable thermodynamic properties as calculated from MD simulations.[71] These restricted 3D pharmacophores also should be more likely to retrieve entropically favorable ligands. The HSRP approach was evaluated for three pharmaceutically relevant target proteins, showing a successful reduction of pharmacophore feature space with a simultaneous decrease in required computing power.

SILCS-Pharm (2014). SILCS-Pharm exploits binding hotspots of probe molecules in MD simulations for 3D pharmacophore generation.[72,73] The SILCS (site identification by ligand competitive saturation) method is employed to sample the surface of proteins in MD simulations with different probe molecules reassembling properties known from pharmacophore features, for example, benzene carbons for aromatic features.[94] The resulting probability maps of the different probe molecules (FragMaps) are Boltzmann-transformed into free energy representations. These free energy FragMaps are finally converted to pharmacophore features and the associated free energies can be used to prioritize feature selection in 3D pharmacophore model generation. The authors showed that 3D pharmacophores generated with SILCS-Pharm often perform better than various docking approaches and 3D pharmacophores generated with the HSRP method described above. SILCS-Pharm was already employed in guiding binding pose predictions of novel inhibitors targeting the oncoproteins Mcl-1 and Bcl-xL.[95]

Dynophores (*dynamic pharmacophores*) (*2015*). Contrary to approaches that gather pharmacophore information from ensembles of different protein conformations based on MD, the dynophore app represents a fully automated implementation of chemical feature-based interaction patterns with MD-based conformational sampling.[66,74] Dynophores (dynamic pharmacophores) sequentially extract interaction points (such as hydrogen bonds, charges, or lipophilic contacts) from each frame of a trajectory according to the ligand atoms involved and their feature type. The resulting super-features can be statistically characterized by occurrence frequency and interaction patterns with the

TABLE 2   Advanced approaches employing the 3D pharmacophore concept

Category	Approach	Description
MD integration	Hydration-site restricted pharmacophore[71]	Thermodynamic properties of water molecules are used to reduce the number of features in apo-based 3D pharmacophore models
	SILCS-Pharm[72,73]	Binding hot spots of probe molecules in MD simulations are exploited for 3D pharmacophore generation
	Dynophore[74]	Fully automated combination of 3D pharmacophores and MD simulations with statistics on spatiotemporal feature occurrency
	Common hits approach[75]	3D pharmacophore models from MD simulations are grouped according to interaction pattern and used for parallel screening
	MYSHAPE[76]	Interaction patterns in MD simulations are analyzed to refine shared-feature pharmacophores
	GRAIL[77]	Molecular interaction fields are abstracted to the pharmacophoric level and averaged over an MD simulation
	Water pharmacophore[78]	Water thermodynamics, docking, and molecular interaction fields are used to generate a single 3D pharmacophore model
	PyRod[79]	Protein environment of water molecules is analyzed to generate dynamic molecular interaction fields for visualization and 3D pharmacophores for virtual screening
	AutoDock Bias[80]	Cosolvent simulations are used to bias a docking algorithm
	Pharmmaker[81]	Cosolvent simulations are analyzed to generate 3D pharmacophores for virtual screening
Machine learning	HSPharm[30]	Random forest decision tree is trained on pharmacophoric fingerprints to reduce the number of features in apo-based 3D pharmacophore models
	PharmIF[82]	Support vector machine is trained on pharmacophoric fingerprints to rank docking poses
	DeepSite and related softwares[83–86]	Convolutional neural network is trained on pharmacophoric descriptors to detect cavities, predict binding affinities, and to design new molecules
Web applications	PharmaGist[87]	Ligand-based 3D pharmacophore generation
	PharmMapper[88]	Target prediction of small molecules employing a database of 3D pharmacophores
	Pharmer related applications[89–91]	3D pharmacophore generation and virtual screening of small molecule databases

protein. Three-dimensional volumetric feature density clouds provide information about the spatial distribution of interactions and barcode plots show the feature occurrence in a time-resolved manner. The dynophore app was implemented within the ilib/LigandScout framework[11] and addresses two shortcomings of classical 3D pharmacophore models: their static character and the geometric simplification of the features.

The application of dynophores proved essential in several studies for rationally explaining phenomena which could not have been unveiled by classical 3D pharmacophores alone. Dynophores were first used to explain an activity cliff of two ligands at the M_2 receptor, which differ in structure by a single double bond (dihydroisoxazole vs. isoxazole moiety) and therefore show the same static 3D pharmacophore.[66] Based on the occurrence frequency and the respective geometric properties of a hydrogen bond acceptor, different strengths of the resulting hydrogen bond between the two ligands could be rationalized. In another example, dynophores were able to unveil a mechanism to overcome drug resistance for HIV-1 reverse transcriptase (RT). In this study, the RT inhibitor rilpivirine was shown to bypass resistance mutations by interacting with alternate residues, stabilizing the inhibitor in the binding pocket.[96] In a virtual screening campaign against the metalloenzyme arginase, dynophores were employed to explore the plasticity of the binding pocket in the presence of small molecule inhibitors and suggested the possibility for additional lipophilic contacts. This resulted in two novel fragment arginase inhibitors that could aid the development of anticancer drugs.[97] The dynophore methodology enables researchers to escape the static nature of classical 3D pharmacophore approaches and provides
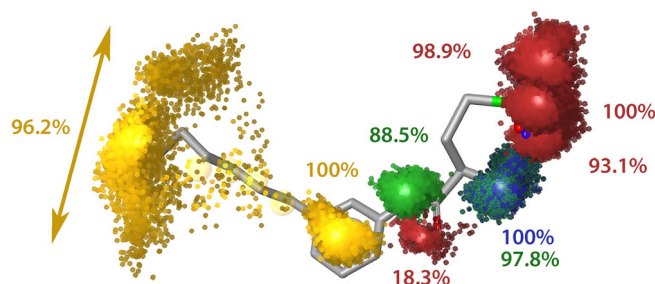
FIGURE 4 Dynophores (dynamic pharmacophores) unveil dynamic binding mode changes of the sphingosine-1-phosphate receptor ligand ML056 during a 100 ns MD simulation. The yellow point clouds indicate lipophilic contacts, the red and green features represent hydrogen bond acceptor or donor, respectively, and a positively charged area is shown as a blue point cloud. The percentages next to the features refer to their occurrence frequency during the simulation. In the example shown, a major part of the molecule remains in its initial orientation resulting in nearly sphere-like distributions of the according feature point clouds (right part). The lipophilic tail is much more flexible within the binding site as indicated by the banana-shaped feature cloud (left part). MD, molecular dynamics

new opportunities to describe and analyze ligand binding modes as dynamic events (Figure 4).[47,66,96–101] The probability density functions representing the feature distribution of the super-features can be directly used for virtual screening, providing new possibilities for efficiently incorporating information from molecular dynamics simulations into fast and efficient virtual screening.

Common hits approach and MYSHAPE (2017). These two approaches select and optimize 3D pharmacophores based on MD simulations. The Common Hits Approach (CHA) groups 3D pharmacophore models obtained from protein–ligand conformations of MD simulations according to their interaction pattern.[75] Representative 3D pharmacophores are subsequently used for computationally costly virtual screening. CHA was retrospectively evaluated on 40 protein–ligand systems, showing improved virtual screening performance for many of the protein–ligand complexes compared to the use of single 3D pharmacophore models. In contrast, MYSHAPE optimizes shared-feature pharmacophores by focusing on pharmacophore features observed during MD simulations of different protein–ligand complexes. This approach was found in a retrospective evaluation against PPARα to perform better than 3D pharmacophores derived from X-ray structure.[76]

GRAIL (2018). The GRids of phArmacophore Interaction fieLds (GRAIL) approach depicts MIFs on the pharmacophore level in MD simulations.[77] Beside pharmacophoric interaction fields, this approach generates information on atom densities for protein, water, and ligand if present. GRAIL was applied to MD simulations of heat shock protein 90 showing that the pharmacophoric interaction fields can contribute to the understanding of the structure–activity relationship of a complexed ligand series.

Water pharmacophore (2018). The Water Pharmacophore (WP) method aims at generating 3D pharmacophores based on thermodynamic properties of hydration sites, similar to the HSPR approach described above.[71,78] WPs are generated for hydration sites by a combination of thermodynamic analysis, MIFs, and docking-based strategies. However, in contrast to HSPR, the WP method generates a single 3D pharmacophore in a highly automated fashion with a comparably low number of involved features granting high performance in virtual screening campaigns. After optimizing parameters for the 3D pharmacophore generation procedure against seven pharmaceutically relevant targets, the authors were able to generate successful 3D pharmacophores for four out of seven targets.

PyRod (2019). Similar to WP and HSRP, the free and open-source software PyRod focuses on water molecules in protein binding pockets to generate 3D pharmacophores for virtual screening.[79] However, instead of determining thermodynamic properties of hydration sites, PyRod analyzes the protein environment of water molecules in protein binding pockets based on fast-to-calculate pharmacophore inspired heuristic scoring functions. This information is further processed to visualize pharmacophoric binding pocket characteristics in the form of dynamic dMIFs and to generate pharmacophore features for virtual screening (Figure 5). Since scoring is only performed in the presence of water molecules, pharmacophore features are preferentially placed at hydration sites with high water occupancy. Replacing such water molecules with a ligand moiety results in a gain of entropy and hence increases the chance for discovery of high affinity ligands. Similar to other apo-based pharmacophore methods, the unrefined 3D pharmacophore contains too many features for efficient virtual screening. Hence, the user must first preselect pharmacophore features based on dMIFs and their arrangement in the binding pocket. The features of this focused 3D pharmacophore are subsequently combined to generate a pharmacophore library based on user-defined characteristics, for example, maximal and
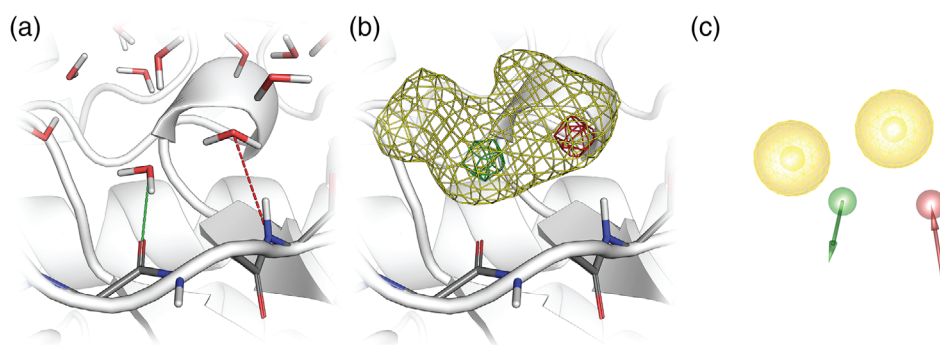
FIGURE 5 PyRod applied to the binding pocket of cyclin-dependent kinase 2. (a) The protein environment of water molecules is analyzed to generate (b) dynamic molecular interaction fields (dMIFs) describing the pharmacophoric characteristics of the binding pocket, (c) which can be translated into pharmacophore features for virtual screening. Yellow—hydrophobic contact, green—hydrogen bond donor, red—hydrogen bond acceptor

minimal number of independent features or hydrophobic contacts. A Python script is provided allowing for evaluation of the generated pharmacophore library using LigandScout.[11] PyRod pharmacophores performed better than docking for three out of five pharmaceutically relevant targets, according to ROC analysis in a retrospective evaluation and provides a directly usable workflow for efficient virtual screening.

AutoDock Bias with Solvent Sites (2019) employs cosolvent based pharmacophores to bias docking algorithms toward hotspots of probe molecule binding for improving virtual screening performance.[80,102,103] First, the authors performed cosolvent MD simulations of proteins in the presence of water and ethanol. Next, the trajectories were analyzed to identify hotspots for binding of the ethanol hydroxyl- and methyl-group. Finally, the calculated free energies of ethanol hotspots were used to introduce an additional energy term to the docking algorithm of AutoDock 4.[28,80] The biased docking performance was retrospectively evaluated, showing improved performance compared to the standard docking procedure in the majority of investigated test systems.[103]

Pharmmaker (2020) analyzes cosolvent simulations to generate 3D Pharmacophores for virtual screening.[81] In the presented case study, cosolvent simulations were performed using six different probe molecules and subsequently employed to assess the druggability of different binding sites with DruGUI.[104] The most druggable binding site was subsequently processed with Pharmmaker by selecting protein residues with high probe-specific affinities and by identifying snapshots that show the most frequent interactions between protein residues and probe molecules. Finally, 3D pharmacophores were generated from selected snapshots and employed for virtual screening with Pharmit,[89] a web application for pharmacophore screening described later in this review.

In addition to 3D pharmacophore modeling, the benefit of multiple protein conformations in enhancing the performance of molecular docking is also frequently discussed.[105] However, ensemble docking in multiple X-ray structures revealed minor improvements in screening performance in only some test systems, which hardly justifies the increased computational costs.[106] Clustering protein conformations could be a solution to reduce the computational costs, but fails to identify the most relevant protein conformations from MD simulations.[107] In contrast to molecular docking, the presented 3D pharmacophore approaches show clear advantages over traditional static pharmacophore modeling. Dynophores, for example, provide a statistical characterization of interactions in the form of volumetric feature density clouds. Escaping from the spherical nature of traditional pharmacophore features in conjunction with the representation of pharmacophore features as probability density functions represents a strong opportunity to boost virtual screening performance. Furthermore, approaches like PyRod consider entropic contribution of ligand binding which can only be poorly estimated by static methods. In comparison to these advanced pharmacophore approaches, ensemble docking represents a computationally expensive parallelization of the same underlying algorithms. Hence, important information contained in MD simulations is not properly considered, resulting in comparably poor improvement of performance.

4.2 | Training machine learning models with pharmacophoric descriptors

In recent years, machine learning and artificial intelligence have witnessed tremendous amount of attention in the public media. The simultaneous improvement in computing power and increase of available data have heavily influenced

the modern drug discovery process.[108] Moreover, the concept of 3D pharmacophores was employed to develop several new machine learning methods.

HS-pharm (*2008*). The Hot-Spots-Guided Receptor-Based Pharmacophores (HS-Pharm) approach trains machine learning models to reduce the number of features from apo-based 3D pharmacophore models.[30] The binding cavities of ~3,500 resolved protein–ligand complexes were analyzed and over 600 k atoms were distributed into interacting and noninteracting groups. Atom-based cavity fingerprints were generated from the gathered cavity atoms, collecting data about pharmacophoric and torsional properties, involved residues, and their protein environment. Decision trees and Bayesian classifiers were trained and tested on these fingerprints to detect cavity atoms important for ligand binding. Evaluation of the approach identified random forest decision trees to perform best according to enrichment and ROC analysis. Finally, this approach was applied to three pharmaceutical relevant drug targets resulting in the generation of 3D pharmacophores performing better than docking in two out of three cases.

Pharm-IF (*2010*). The pharmacophore-based interaction fingerprint Pharm-IF was evaluated as input for several machine learning algorithms to rank docking poses of small molecules.[82] Interaction fingerprints encoding the type and distance of interaction partners were generated for all available atomistic models of five pharmaceutically relevant drug targets. These were subsequently used to train several machine learning algorithms to rank docking poses of known actives and decoys. In a retrospective evaluation, Pharm-IF fingerprints in combination with support-vector machines showed the best enrichment, outperforming other machine learning algorithms and a docking scoring function. Employing Pharm-IF to train machine learning algorithms resulted in better enrichment compared to employing PLIF, a protein–ligand interaction fingerprint implemented in MOE.[13] In contrast to Pharm-IF, this fingerprint does not encode the distance, which suggests an important contribution of the distance to successful predictions. Interestingly, learning on more than five crystal structures enabled models to predict activity better than docking scores for all studied targets.

DeepSite and related softwares (*2017*). The DeepSite software employs convolutional neural networks typically used for analyzing visual imagery to predict the druggability of protein binding pockets.[83] Atomic-based pharmacophoric descriptors were assigned to grid points covering the protein of approximately 7,000 protein–ligand complexes. These grids were subsequently divided into subgrids, which were labeled as a binding site if their geometric center was within 4 Å of the binding site's geometric center. These 3D images of the binding pocket represented by 3D grids of pharmacophoric descriptors were used to train a convolutional neural network. The DeepSite cavity detection was found to perform better than other state-of-the-art detection algorithms. Similar approaches that also use pharmacophoric descriptors to train convolutional neural networks were later used to predict binding affinities of small molecules (K_{DEEP}[84]) and to guide the design of novel molecules (LigVoxel,[85] LigDream[86]). All of the aforementioned approaches are implemented in software packages for local installation but can also be used free of charge as web applications.

4.3 | Web applications employing 3D pharmacophores

Although web applications do not necessarily represent an advancement of the 3D pharmacophore concept itself, they do advance their usability, bringing 3D pharmacophores into the internet age. Since a local installation of software is not needed and all web applications presented are freely available for academic research, users can circumvent license fees and screen databases with millions of molecules without the need for high-performance computational resources. Thus, 3D pharmacophore searches become available to a larger number of users.

PharmaGist (*2008*). The PharmaGist web application allows for ligand-based 3D pharmacophore generation.[87] Each submitted ligand is analyzed for rotatable bonds important for flexible molecule alignment and for pharmacophoric features used for alignment to a reference ligand. A user-specified maximal number of 3D pharmacophores is generated, and the output comprising 3D pharmacophores and aligned molecules can be downloaded.

PharmMapper (*2010*). The PharmMapper web application can be used for target fishing of small molecules important for off-target prediction and polypharmacology studies.[88] More than 53,000 3D pharmacophore models were generated from approximately 23,000 protein database (PDB) crystal structures based on protein–ligand interactions or involving a cavity detection algorithm to identify potential allosteric binding sites. Submitted molecules are scored by their match to all deposited 3D pharmacophore models.

Pharmer-based web applications (*2012*). The Pharmer virtual screening software[10] was employed for several web applications enabling the efficient virtual screening of several small molecule databases. AnchorQuery is specialized for

the identification of protein–protein interaction inhibitors.[90] The user uploads a protein–protein complex and specifies an anchor residue that is likely to be important for protein–protein interaction. This anchor residue will become part of a 3D pharmacophore that is used to screen a multimillion compound library of synthesizable small molecules. ZINCPharmer can be used to perform a virtual screening of the ZINC database, a free virtual library collecting commercially available compounds from different vendors.[91,109] The web application supports the import of 3D pharmacophores but can also be used to generate 3D pharmacophore models from scratch. Lastly, Pharmit enables the virtual screening of several commercial vendors as well as other noncommercial databases including ChEMBL and PubChem.[89,110,111] The 3D pharmacophores can be either designed from a protein–ligand complex or, a ligand. The import of 3D pharmacophore models in several data formats is also supported.

5 | CONCLUSION

In this review, we have given an overview of the principles of 3D pharmacophores and their role in drug discovery. The fact that 3D pharmacophore models are universal, editable, and comprehensive allows them to be applied in different scenarios.

A major application field is the identification of novel ligands through virtual screening. For this purpose, 3D pharmacophore models are the sole technique that can be applied in either a ligand-based or a structure-based manner. In both ways, 3D pharmacophore models are computationally very efficient, enabling the virtual screening of very large databases. The basic concept of abstracting chemical functionality allows for scaffold hopping and enriches the chemical diversity of hit lists. Altogether, this grants researchers more flexibility regarding available data, computational resources, and testing capabilities. The case studies that we selected highlight the power of pharmacophore-based virtual screening for drug discovery and show their applicability to challenging targets. Also, increasingly popular fragment-based drug discovery campaigns can benefit from pharmacophore screening by a dramatic reduction of fragments tested in vitro and by rationalizing fragment growing with constant fragment core interactions.[54,55]

Besides virtual screening, 3D pharmacophores are well suited to study and visualize binding modes of drug-like molecules. Their composition of a limited number of chemically defined interaction features make them understandable and intuitive. This represents a major advantage in interdisciplinary projects, since 3D pharmacophore models are able to rationalize various pharmacological effects. For this objective, 3D pharmacophores are typically combined with other methods such as docking, MD simulations, or machine learning. The selected case studies for this field underline the power of 3D pharmacophores to mechanistically explain and understand protein functionality. Additionally, 3D pharmacophores are an excellent tool for communication between researchers, a factor that is often underestimated.

However, besides the aforementioned advantages and possibilities, classic 3D pharmacophore models also have certain drawbacks. They represent static models for highly dynamic systems and their interaction features are restricted to simple geometries (e.g., spherical features). Moreover, they share a shortcoming with other modeling techniques, which all are focused on estimating the enthalpy of molecular interactions but are suboptimal for the description of entropic effects. However, enthalpy and entropy both contribute to the change in free energy of ligand binding to a macromolecule. Although the basic concept of 3D pharmacophore generation and its application to virtual screening has not changed in the last 30 years, there are various developments in the field that aim at addressing these shortcomings.

The combination of 3D pharmacophore models with MDs is therefore a consequent evolution with great potential. Different approaches to integrating MDs into 3D pharmacophore modeling have been reported and described in this review.[71–75,77–79,81,103] However, only the dynophore method represents a fully automated approach, which tackles two drawbacks of classical 3D pharmacophores at a time.[74] The dynophore application reveals a new perspective on ligand binding by providing visualization of pharmacophoric features that escape from the traditional spherical geometry and by delivering statistics that report feature occurrence frequencies and different binding modes over the course of a trajectory. The direct usage of these property-density functions for virtual screening would represent a true paradigm shift in 3D pharmacophore modeling.

Several advanced approaches also consider entropic effects of ligand binding for 3D pharmacophore modeling.[71–73,78,79,81,103] PyRod, for instance, analyzes the protein environment of water molecules in MD simulations, which allows for placement of pharmacophore features at hydration sites with certain thermodynamic characteristics.[79] Such hydration sites may harbor water molecules in a highly hydrophobic protein environment or heavily restrain water molecules via hydrogen bonds and the shape of the binding pocket. The restriction of 3D pharmacophores to entropically and enthalpically important sites render such approaches valuable tools for virtual screening campaigns,

especially for those generating 3D pharmacophores from an apo structure. Importantly, PyRod is a free and open-access tool making such strategies accessible to a broader user base.

The combination of 3D pharmacophore concept and machine learning/artificial intelligence is only in its beginning stages. Although some approaches already exist,[30,82–86] we predict an increasing number of studies and methods that aim to use pharmacophore features as descriptors or try to generate 3D pharmacophores from big data. Another trend that we observe is the availability of freely available web services for pharmacophore-based virtual screening.[87–91]

The recent developments in the field of 3D pharmacophores are promising and afford the opportunity to employ 3D pharmacophores in ever-increasing ways and more challenging situations, such as multitarget prediction, modeling binding kinetics, or pathway-specific receptor activation. Overall, 3D pharmacophores represent an essential part of the toolbox for computer-aided drug design and are perfectly apt to identify novel ligands and understand their interaction with the macromolecular target.

CONFLICT OF INTEREST
The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS
David Schaller: Conceptualization; visualization; writing-original draft; writing-review and editing. **Dora Sribar:** Visualization; writing-original draft; writing-review and editing. **Tessa Noonan:** Writing-original draft; writing-review and editing. **Lihua Deng:** Writing-original draft; writing-review and editing. **Trung Ngoc Nguyen:** Writing-original draft; writing-review and editing. **Szymon Pach:** Writing-original draft; writing-review and editing. **David Machalz:** Writing-original draft; writing-review and editing. **Marcel Bermudez:** Conceptualization; visualization; writing-original draft; writing-review and editing. **Gerhard Wolber:** Conceptualization; supervision; visualization; writing-original draft; writing-review and editing.

ORCID
David Schaller https://orcid.org/0000-0002-1881-4518
Theresa Noonan https://orcid.org/0000-0003-4924-6848
Lihua Deng https://orcid.org/0000-0002-7173-6897
Trung Ngoc Nguyen https://orcid.org/0000-0002-7415-4390
Szymon Pach https://orcid.org/0000-0001-6109-7123
David Machalz https://orcid.org/0000-0001-8634-9411
Marcel Bermudez https://orcid.org/0000-0002-7421-3282
Gerhard Wolber https://orcid.org/0000-0002-5344-0048

RELATED WIREs ARTICLES
Computational chemogenomics
Computational methods for scaffold hopping
Generation of three-dimensional pharmacophore models
Computational close up on protein-protein interactions: how to unravel the invisible using molecular dynamics simulations?

REFERENCES
1. Güner OF, Bowen JP. Setting the record straight: The origin of the pharmacophore concept. *J Chem Inf Model*. 2014;54(5):1269–1283. https://doi.org/10.1021/ci5000533
2. Reymond J-L, van Deursen R, Blum LC, Ruddigkeit L. Chemical space as a source for new drugs. *Med Chem Commun*. 2010;1(1):30. https://doi.org/10.1039/c0md00020e
3. Gromski PS, Henson AB, Granda JM, Cronin L. How to explore chemical space using algorithms and automation. *Nat Rev Chem*. 2019;3(2):119–128. https://doi.org/10.1038/s41570-018-0066-y

4. Van Drie JH. Generation of three-dimensional pharmacophore models. *Wiley Interdis Rev*. 2013;*3*(5):449–464. https://doi.org/10.1002/wcms.1129

5. Voth AR, Khuu P, Oishi K, Ho PS. Halogen bonds as orthogonal molecular interactions to hydrogen bonds. *Nat Chem*. 2009;*1*(1): 74–79. https://doi.org/10.1038/nchem.112

6. Lu Y, Shi T, Wang Y, et al. Halogen bonding-a novel interaction for rational drug design? *J Med Chem*. 2009;*52*(9):2854–2862. https://doi.org/10.1021/jm9000133

7. Wilcken R, Zimmermann MO, Lange A, Joerger AC, Boeckler FM. Principles and applications of halogen bonding in medicinal chemistry and chemical biology. *J Med Chem*. 2013;*56*(4):1363–1388. https://doi.org/10.1021/jm3012068

8. Leach AR, Gillet VJ, Lewis RA, Taylor R. Three-dimensional pharmacophore methods in drug discovery. *J Med Chem*. 2010;*53*(2): 539–558. https://doi.org/10.1021/jm900817u

9. Baroni M, Cruciani G, Sciabola S, Perruccio F, Mason JS. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for ligands and proteins (FLAP): Theory and application. *J Chem Inf Model*. 2007;*47*(2):279–294. https://doi.org/10.1021/ci600253e

10. Koes DR, Camacho CJ. Pharmer: Efficient and exact pharmacophore search. *J Chem Inf Model*. 2011;*51*(6):1307–1314. https://doi.org/10.1021/ci200097m

11. Wolber G, Langer T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model*. 2005;*45*(1):160–169. https://doi.org/10.1021/ci049885e

12. Barnum D, Greene J, Smellie A, Sprague P. Identification of common functional configurations among molecules. *J Chem Inf Comput Sci*. 1996;*36*(3):563–571. https://doi.org/10.1021/ci950273r

13. Chemical Computing Group. Molecular operating environment (MOE). Montreal, QC, Canada; 2010.

14. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA. PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des*. 2006;*20*(10–11): 647–671. https://doi.org/10.1007/s10822-006-9087-6

15. Taminau J, Thijs G, De Winter H. Pharao: Pharmacophore alignment and optimization. *J Mol Graph Model*. 2008;*27*(2):161–169. https://doi.org/10.1016/j.jmgm.2008.04.003

16. Certara. UNITY, SYBYL-X. St. Louis, MO; 2013.

17. Cheeseright TJ, Mackey MD, Scoffin RA. High content pharmacophores from molecular fields: A biologically relevant method for comparing and understanding ligands. *Curr Comput Aided Drug Des*. 2011;*7*(3):190–205. https://doi.org/10.2174/157340911796504314

18. Schuster D, Wolber G. Identification of bioactive natural products by pharmacophore-based virtual screening. *Curr Pharm Des*. 2010;*16* (15):1666–1681. https://doi.org/10.2174/138161210791164072

19. Wolber G, Dornhofer AA, Langer T. Efficient overlay of small organic molecules using 3D pharmacophores. *J Comput Aided Mol Des*. 2006;*20*(12):773–788. https://doi.org/10.1007/s10822-006-9078-7

20. Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem*. 1994;*37*(24):4130–4146. https://doi.org/10.1021/jm00050a010

21. Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc*. 1988;*110*(18):5959–5967. https://doi.org/10.1021/ja00226a005

22. Friedrich N-O, de Bruyn KC, Flachsenberg F, Sommer K, Rarey M, Kirchmair J. Benchmarking commercial conformer ensemble generators. *J Chem Inf Model*. 2017;*57*(11):2719–2728. https://doi.org/10.1021/acs.jcim.7b00505

23. Spitzer GM, Wellenzohn B, Laggner C, Langer T, Liedl KR. DNA minor groove pharmacophores describing sequence specific properties. *J Chem Inf Model*. 2007;*47*(4):1580–1589. https://doi.org/10.1021/ci600500v

24. Friesner RA, Murphy RB, Repasky MP, et al. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem*. 2006;*49*(21):6177–6196. https://doi.org/10.1021/jm051256o

25. Salam NK, Nuti R, Sherman W. Novel method for generating structure-based pharmacophores using energetic analysis. *J Chem Inf Model*. 2009;*49*(10):2356–2368. https://doi.org/10.1021/ci900212v

26. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem*. 1985;*28*(7):849–857. https://doi.org/10.1021/jm00145a002

27. von Itzstein M, Wu W-Y, Kok GB, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*. 1993; *363*(6428):418–423. https://doi.org/10.1038/363418a0

28. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009;*30*(16):2785–2791. https://doi.org/10.1002/jcc.21256

29. Mortier J, Dhakal P, Volkamer A. Truly target-focused pharmacophore modeling: A novel tool for mapping intermolecular surfaces. *Molecules*. 2018;*23*(8):1959. https://doi.org/10.3390/molecules23081959

30. Barillari C, Marcou G, Rognan D. Hot-spots-guided receptor-based pharmacophores (HS-pharm): A knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *J Chem Inf Model*. 2008;*48*(7): 1396–1410. https://doi.org/10.1021/ci800064z

31. Hurst T. Flexible 3D searching: The directed tweak technique. *J Chem Inf Model*. 1994;*34*(1):190–196. https://doi.org/10.1021/ci00017a025

32. Seidel T, Ibis G, Bendix F, Wolber G. Strategies for 3D pharmacophore-based virtual screening. *Drug Discov Today Technol*. 2010;*7*(4): e221–e228. https://doi.org/10.1016/j.ddtec.2010.11.004

33. Hawkins PCD. Conformation generation: The state of the art. *J Chem Inf Model*. 2017;57(8):1747–1756. https://doi.org/10.1021/acs.jcim.7b00221

34. Kirchmair J, Distinto S, Markt P, et al. How to optimize shape-based virtual screening: Choosing the right query and including chemical information. *J Chem Inf Model*. 2009;49(3):678–692. https://doi.org/10.1021/ci8004226

35. Kirchmair J, Distinto S, Schuster D, Spitzer G, Langer T, Wolber G. Enhancing drug discovery through in silico screening: Strategies to increase true positives retrieval rates. *Curr Med Chem*. 2008;15(20):2040–2053. https://doi.org/10.2174/092986708785132843

36. Kirchmair J, Wolber G, Laggner C, Langer T. Comparative performance assessment of the conformational model generators omega and catalyst: A large-scale survey on the retrieval of protein-bound ligand conformations. *J Chem Inf Model*. 2006;46(4):1848–1861. https://doi.org/10.1021/ci060084g

37. Sanders MPA, Barbosa AJM, Zarzycka B, et al. Comparative analysis of pharmacophore screening tools. *J Chem Inf Model*. 2012;52(6):1607–1620. https://doi.org/10.1021/ci2005274

38. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J Med Chem*. 2012;55(14):6582–6594. https://doi.org/10.1021/jm300687e

39. Ntie-Kang F, Simoben CV, Karaman B, et al. Pharmacophore modeling and in silico toxicity assessment of potential anticancer agents from African medicinal plants. *Drug Des Devel Ther*. 2016;10:2137–2154. https://doi.org/10.2147/DDDT.S108118

40. Braga RC, Andrade CH. Assessing the performance of 3D pharmacophore models in virtual screening: How good are they? *Curr Top Med Chem*. 2013;13(9):1127–1138. https://doi.org/10.2174/1568026611313090010

41. Triballeau N, Acher F, Brabet I, Pin J-P, Bertrand H-O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem*. 2005;48(7):2534–2547. https://doi.org/10.1021/jm049092j

42. Kumar H, Kawai T, Akira S. Toll-like receptors and innate immunity. *Biochem Biophys Res Commun*. 2009;388(4):621–625. https://doi.org/10.1016/j.bbrc.2009.08.062

43. Murgueitio MS, Rakers C, Frank A, Wolber G. Balancing inflammation: Computational design of small-molecule toll-like receptor modulators. *Trends Pharmacol Sci*. 2017;38(2):155–168. https://doi.org/10.1016/j.tips.2016.10.007

44. Murgueitio MS, Henneke P, Glossmann H, Santos-Sierra S, Wolber G. Prospective virtual screening in a sparse data scenario: Design of small-molecule TLR2 antagonists. *ChemMedChem*. 2014;9(4):813–822. https://doi.org/10.1002/cmdc.201300445

45. Grabowski M, Murgueitio MS, Bermudez M, Rademann J, Wolber G, Weindl G. Identification of a pyrogallol derivative as a potent and selective human TLR2 antagonist by structure-based virtual screening. *Biochem Pharmacol*. 2018;154:148–160. https://doi.org/10.1016/j.bcp.2018.04.018

46. Grabowski M, Murgueitio MS, Bermudez M, Wolber G, Weindl G. The novel small-molecule antagonist MMG-11 preferentially inhibits TLR2/1 signaling. *Biochem Pharmacol*. 2020;171:113687. https://doi.org/10.1016/j.bcp.2019.113687

47. Šribar D, Grabowski M, Murgueitio MS, Bermudez M, Weindl G, Wolber G. Identification and characterization of a novel chemotype for human TLR8 inhibitors. *Eur J Med Chem*. 2019;179:744–752. https://doi.org/10.1016/j.ejmech.2019.06.084

48. Singh J, Petter RC, Baillie TA, Whitty A. The resurgence of covalent drugs. *Nat Rev Drug Discov*. 2011;10(4):307–317. https://doi.org/10.1038/nrd3410

49. De Cesco S, Kurian J, Dufresne C, Mittermaier AK, Moitessier N. Covalent inhibitors design and discovery. *Eur J Med Chem*. 2017;138:96–114. https://doi.org/10.1016/j.ejmech.2017.06.019

50. Johnson DS, Weerapana E, Cravatt BF. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med Chem*. 2010;2(6):949–964. https://doi.org/10.4155/fmc.10.21

51. Bauer RA. Covalent inhibitors in drug discovery: From accidental discoveries to avoided liabilities and designed therapies. *Drug Discov Today*. 2015;20(9):1061–1073. https://doi.org/10.1016/j.drudis.2015.05.005

52. London N, Miller RM, Krishnan S, et al. Covalent docking of large libraries for the discovery of chemical probes. *Nat Chem Biol*. 2014;10(12):1066–1072. https://doi.org/10.1038/nchembio.1666

53. Lonsdale R, Burgess J, Colclough N, et al. Expanding the armory: Predicting and tuning covalent warhead reactivity. *J Chem Inf Model*. 2017;57(12):3124–3137. https://doi.org/10.1021/acs.jcim.7b00553

54. Schulz R, Atef A, Becker D, et al. Phenylthiomethyl ketone-based fragments show selective and irreversible inhibition of enteroviral 3C proteases. *J Med Chem*. 2018;61(3):1218–1230. https://doi.org/10.1021/acs.jmedchem.7b01440

55. Murray CW, Rees DC. The rise of fragment-based drug discovery. *Nat Chem*. 2009;1(3):187–192. https://doi.org/10.1038/nchem.217

56. Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE. Trends in GPCR drug discovery: New agents, targets and indications. *Nat Rev Drug Discov*. 2017;16(12):829–842. https://doi.org/10.1038/nrd.2017.178

57. Frandsen IO, Boesgaard MW, Fidom K, et al. Identification of histamine H3 receptor ligands using a new crystal structure fragment-based method. *Sci Rep*. 2017;7(1):4829. https://doi.org/10.1038/s41598-017-05058-w

58. Fidom K, Isberg V, Hauser AS, et al. A new crystal structure fragment-based pharmacophore method for G protein-coupled receptors. *Methods*. 2015;71:104–112. https://doi.org/10.1016/j.ymeth.2014.09.009

59. Schaller D, Hagenow S, Stark H, Wolber G. Ligand-guided homology modeling drives identification of novel histamine H3 receptor ligands. *PLoS One*. 2019;14(6):e0218820. https://doi.org/10.1371/journal.pone.0218820

60. Temml V, Garscha U, Romp E, et al. Discovery of the first dual inhibitor of the 5-lipoxygenase-activating protein and soluble epoxide hydrolase using pharmacophore-based virtual screening. *Sci Rep*. 2017;7(1):42751. https://doi.org/10.1038/srep42751

61. Waltenberger B, Garscha U, Temml V, et al. Discovery of potent soluble epoxide hydrolase (sEH) inhibitors by pharmacophore-based virtual screening. *J Chem Inf Model*. 2016;*56*(4):747–762. https://doi.org/10.1021/acs.jcim.5b00592

62. Yao T-T, Fang S-W, Li Z-S, et al. Discovery of novel succinate dehydrogenase inhibitors by the integration of in silico library design and pharmacophore mapping. *J Agric Food Chem*. 2017;*65*(15):3204–3211. https://doi.org/10.1021/acs.jafc.7b00249

63. Rakers C, Schumacher F, Meinl W, Glatt H, Kleuser B, Wolber G. In silico prediction of human sulfotransferase 1E1 activity guided by pharmacophores from molecular dynamics simulations. *J Biol Chem*. 2016;*291*(1):58–71. https://doi.org/10.1074/jbc. M115.685610

64. Tkachenko A, Bermudez M, Irmer-Stooff S, et al. Nuclear transport of the human aryl hydrocarbon receptor and subsequent gene induction relies on its residue histidine 291. *Arch Toxicol*. 2018;*92*(3):1151–1160. https://doi.org/10.1007/s00204-017-2129-0

65. Bermudez M, Rakers C, Wolber G. Structural characteristics of the allosteric binding site represent a key to subtype selective modulators of muscarinic acetylcholine receptors. *Mol Inform*. 2015;*34*(8):526–530. https://doi.org/10.1002/minf.201500025

66. Bock A, Bermudez M, Krebs F, et al. Ligand binding ensembles determine graded agonist efficacies at a G protein-coupled receptor. *J Biol Chem*. 2016;*291*(31):16375–16389. https://doi.org/10.1074/jbc.M116.735431

67. Bermudez M, Bock A, Krebs F, et al. Ligand-specific restriction of extracellular conformational dynamics constrains signaling of the M2 muscarinic receptor. *ACS Chem Biol*. 2017;*12*(7):1743–1748. https://doi.org/10.1021/acschembio.7b00275

68. Bermudez M, Bock A. Does divergent binding pocket closure drive ligand bias for class a GPCRs? *Trends Pharmacol Sci*. 2019;*40*(4): 236–239. https://doi.org/10.1016/j.tips.2019.02.005

69. Bermudez M, Nguyen TN, Omieczynski C, Wolber G. Strategies for the discovery of biased GPCR ligands. *Drug Discov Today*. 2019;*24* (4):1031–1037. https://doi.org/10.1016/j.drudis.2019.02.010

70. Agnetta L, Bermudez M, Riefolo F, et al. Fluorination of photoswitchable muscarinic agonists tunes receptor pharmacology and photochromic properties. *J Med Chem*. 2019;*62*(6):3009–3020. https://doi.org/10.1021/acs.jmedchem.8b01822

71. Hu B, Lill MA. Protein pharmacophore selection using hydration-site analysis. *J Chem Inf Model*. 2012;*52*(4):1046–1060. https://doi. org/10.1021/ci200620h

72. Yu W, Lakkaraju SK, Raman EP, MacKerell AD. Site-identification by ligand competitive saturation (SILCS) assisted pharmacophore modeling. *J Comput Aided Mol Des*. 2014;*28*(5):491–507. https://doi.org/10.1007/s10822-014-9728-0

73. Yu W, Lakkaraju SK, Raman EP, Fang L, MacKerell AD. Pharmacophore modeling using site-identification by ligand competitive saturation (SILCS) with multiple probe molecules. *J Chem Inf Model*. 2015;*55*(2):407–420. https://doi.org/10.1021/ci500691p

74. Sydow D. Dynophores: Novel dynamic pharmacophores (Master Thesis). Humboldt-Universität zu Berlin; 2015.

75. Wieder M, Garon A, Perricone U, et al. Common hits approach: Combining pharmacophore modeling and molecular dynamics simulations. *J Chem Inf Model*. 2017;*57*(2):365–385. https://doi.org/10.1021/acs.jcim.6b00674

76. Perricone U, Wieder M, Seidel T, et al. A molecular dynamics-shared pharmacophore approach to boost early-enrichment virtual screening: A case study on peroxisome proliferator-activated receptor α. *ChemMedChem*. 2017;*12*(16):1399–1407. https://doi.org/10. 1002/cmdc.201600526

77. Schuetz DA, Seidel T, Garon A, et al. GRAIL: Grids of pharmacophore interaction fields. *J Chem Theory Comput*. 2018;*14*(9): 4958–4970. https://doi.org/10.1021/acs.jctc.8b00495

78. Jung SW, Kim M, Ramsey S, Kurtzman T, Cho AE. Water pharmacophore: Designing ligands using molecular dynamics simulations with water. *Sci Rep*. 2018;*8*(1):1–11. https://doi.org/10.1038/s41598-018-28546-z

79. Schaller D, Pach S, Wolber G. PyRod: Tracing water molecules in molecular dynamics simulations. *J Chem Inf Model*. 2019;*59*(6): 2818–2829. https://doi.org/10.1021/acs.jcim.9b00281

80. Arcon JP, Modenutti CP, Avendaño D, et al. AutoDock bias: Improving binding mode prediction and virtual screening using known protein-ligand interactions. Cowen L, editor. *Bioinformatics*. 2019;*35*(19):3836–3838. https://doi.org/10.1093/bioinformatics/btz152

81. Lee JY, Krieger JM, Li H, Bahar I. Pharmmaker: Pharmacophore modeling and hit identification based on druggability simulations. *Protein Sci*. 2020;*29*(1):76–86. https://doi.org/10.1002/pro.3732

82. Sato T, Honma T, Yokoyama S. Combining machine learning and pharmacophore-based interaction fingerprint for in silico screening. *J Chem Inf Model*. 2010;*50*(1):170–185. https://doi.org/10.1021/ci900382e

83. Jiménez J, Doerr S, Martínez-Rosell G, Rose AS, De Fabritiis G. DeepSite: Protein-binding site predictor using 3D-convolutional neural networks. Valencia a, editor. *Bioinformatics*. 2017;*33*(19):3036–3042. https://doi.org/10.1093/bioinformatics/btx350

84. Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G. K DEEP: Protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. *J Chem Inf Model*. 2018;*58*(2):287–296. https://doi.org/10.1021/acs.jcim.7b00650

85. Škalič M, Varela-Rial A, Jiménez J, Martínez-Rosell G, De Fabritiis G. LigVoxel: Inpainting binding pockets using 3D-convolutional neural networks. Valencia a, editor. *Bioinformatics*. 2019;*35*(2):243–250. https://doi.org/10.1093/bioinformatics/bty583

86. Škalič M, Jiménez J, Sabbadin D, De Fabritiis G. Shape-based generative modeling for de novo drug design. *J Chem Inf Model*. 2019;*59* (3):1205–1214. https://doi.org/10.1021/acs.jcim.8b00706

87. Schneidman-Duhovny D, Dror O, Inbar Y, Nussinov R, Wolfson HJ. PharmaGist: A webserver for ligand-based pharmacophore detection. *Nucleic Acids Res*. 2008;*36*:W223–W228. https://doi.org/10.1093/nar/gkn187

88. Wang X, Shen Y, Wang S, et al. PharmMapper 2017 update: A web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Res*. 2017;*45*(W1):W356–W360. https://doi.org/10.1093/nar/gkx374

89. Sunseri J, Koes DR. Pharmit: Interactive exploration of chemical space. *Nucleic Acids Res*. 2016;*44*(W1):W442–W448. https://doi.org/10. 1093/nar/gkw287

90. Koes D, Khoury K, Huang Y, et al. Enabling large-scale design, synthesis and validation of small molecule protein-protein antagonists. *PLoS One*. 2012;*7*(3):e32839. https://doi.org/10.1371/journal.pone.0032839

91. Koes DR, Camacho CJ. ZINCPharmer: Pharmacophore search of the ZINC database. *Nucleic Acids Res*. 2012;*40*(W1):W409–W414. https://doi.org/10.1093/nar/gks378

92. Carlson HA, Masukawa KM, Rubins K, et al. Developing a dynamic pharmacophore model for HIV-1 integrase. *J Med Chem*. 2000;*43* (11):2100–2114. https://doi.org/10.1021/jm990322h

93. Damm KL, Carlson HA. Exploring experimental sources of multiple protein conformations in structure-based drug design. *J Am Chem Soc*. 2007;*129*(26):8225–8235. https://doi.org/10.1021/ja0709728

94. Guvench O, MacKerell AD. Computational fragment-based binding site identification by ligand competitive saturation. Jacobson MP, editor. *PLoS Comput Biol*. 2009;*5*(7):e1000435. https://doi.org/10.1371/journal.pcbi.1000435

95. Lanning ME, Yu W, Yap JL, et al. Structure-based design of N-substituted 1-hydroxy-4-sulfamoyl-2-naphthoates as selective inhibitors of the Mcl-1 oncoprotein. *Eur J Med Chem*. 2016;*113*:273–292. https://doi.org/10.1016/j.ejmech.2016.02.006

96. Nizami B, Sydow D, Wolber G, Honarparvar B. Molecular insight on the binding of NNRTI to K103N mutated HIV-1 RT: Molecular dynamics simulations and dynamic pharmacophore analysis. *Mol Biosyst*. 2016;*12*(11):3385–3395. https://doi.org/10.1039/C6MB00428H

97. Mortier J, Prévost JRC, Sydow D, et al. Arginase structure and inhibition: Catalytic site plasticity reveals new modulation possibilities. *Sci Rep*. 2017;*7*(1):13616. https://doi.org/10.1038/s41598-017-13366-4

98. Bergant K, Janežič M, Valjavec K, et al. Structure-guided optimization of 4,6-substituted-1,3,5-triazin-2(1H)-ones as catalytic inhibitors of human DNA topoisomerase IIα. *Eur J Med Chem*. 2019;*175*:330–348. https://doi.org/10.1016/j.ejmech.2019.04.055

99. Durairaj P, Fan L, Machalz D, Wolber G, Bureik M. Functional characterization and mechanistic modeling of the human cytochrome P450 enzyme CYP4A22. *FEBS Lett*. 2019;*593*(16):2214–2225. https://doi.org/10.1002/1873-3468.13489

100. Naß A, Schaller D, Wolber G. Assessment of flexible shape complementarity: New opportunities to explain and induce selectivity in ligands of protein tyrosine phosphatase 1B. *Mol Inform*. 2019;*38*(5):1800141. https://doi.org/10.1002/minf.201800141

101. Vitorović-Todorović MD, Worek F, Perdih A, Bauk SĐ, Vujatović TB, Cvijetić IN. The in vitro protective effects of the three novel nanomolar reversible inhibitors of human cholinesterases against irreversible inhibition by organophosphorous chemical warfare agents. *Chem Biol Interact*. 2019;*309*:108714. https://doi.org/10.1016/j.cbi.2019.06.027

102. Seco J, Luque FJ, Barril X. Binding site detection and druggability index from first principles. *J Med Chem*. 2009;*52*(8):2363–2371. https://doi.org/10.1021/jm801385d

103. Arcon JP, Defelipe LA, Lopez ED, et al. Cosolvent-based protein pharmacophore for ligand enrichment in virtual screening. *J Chem Inf Model*. 2019;*59*(8):3572–3583. https://doi.org/10.1021/acs.jcim.9b00371

104. Bakan A, Nevins N, Lakdawala AS, Bahar I. Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. *J Chem Theory Comput*. 2012;*8*(7):2435–2447. https://doi.org/10.1021/ct300117j

105. Amaro RE, Baudry J, Chodera J, et al. Ensemble docking in drug discovery. *Biophys J*. 2018;*114*(10):2271–2278. https://doi.org/10.1016/j.bpj.2018.02.038

106. Korb O, Olsson TSG, Bowden SJ, et al. Potential and limitations of ensemble docking. *J Chem Inf Model*. 2012;*52*(5):1262–1274. https://doi.org/10.1021/ci2005934

107. Evangelista Falcon W, Ellingson SR, Smith JC, Baudry J. Ensemble docking in drug discovery: How many protein configurations from molecular dynamics simulations are needed to reproduce known ligand binding? *J Phys Chem B*. 2019;*123*(25):5189–5195. https://doi.org/10.1021/acs.jpcb.8b11491

108. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019;*18*(6):463–477. https://doi.org/10.1038/s41573-019-0024-5

109. Sterling T, Irwin JJ. ZINC 15—Ligand discovery for everyone. *J Chem Inf Model*. 2015;*55*(11):2324–2337. https://doi.org/10.1021/acs.jcim.5b00559

110. Gaulton A, Hersey A, Nowotka M, et al. The ChEMBL database in 2017. *Nucleic Acids Res*. 2017;*45*(D1):D945–D954. https://doi.org/10.1093/nar/gkw1074

111. Kim S, Chen J, Cheng T, et al. PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res*. 2019;*47*(D1):D1102–D1109. https://doi.org/10.1093/nar/gky1033

3.2.4   Molecular Dynamics Simulations

Structure-based drug design and virtual screening campaigns typically rely on crystal structures representing a single conformation of a very dynamic system (Fig 8). Molecular dynamics (MD) simulations provide access to the structural flexibility of the drug-target system making it an increasingly popular tool in the drug discovery process [130, 131].

MD simulation programs like Desmond [132], Amber [133], CHARMM [134] and GROMACS [135] use force fields that describe each atom of the studied system with parameters for (partial) charge, mass and van der Waals radius [136]. The energies of bonded and non-bonded interactions between the atoms of the system can be accessed with potential energy functions. The positions and velocities of atoms over time are calculated by integration of Newton's laws of motions generating a trajectory of successive system configurations [136].

Popular applications of MD simulations in drug design include the local sampling of protein conformations for docking studies, monitoring the association, residence and dissociation of small molecules at their macromolecular target, analyzing the role of water molecules in drug target association and predicting thermodynamic properties for free energy calculations [130, 131].
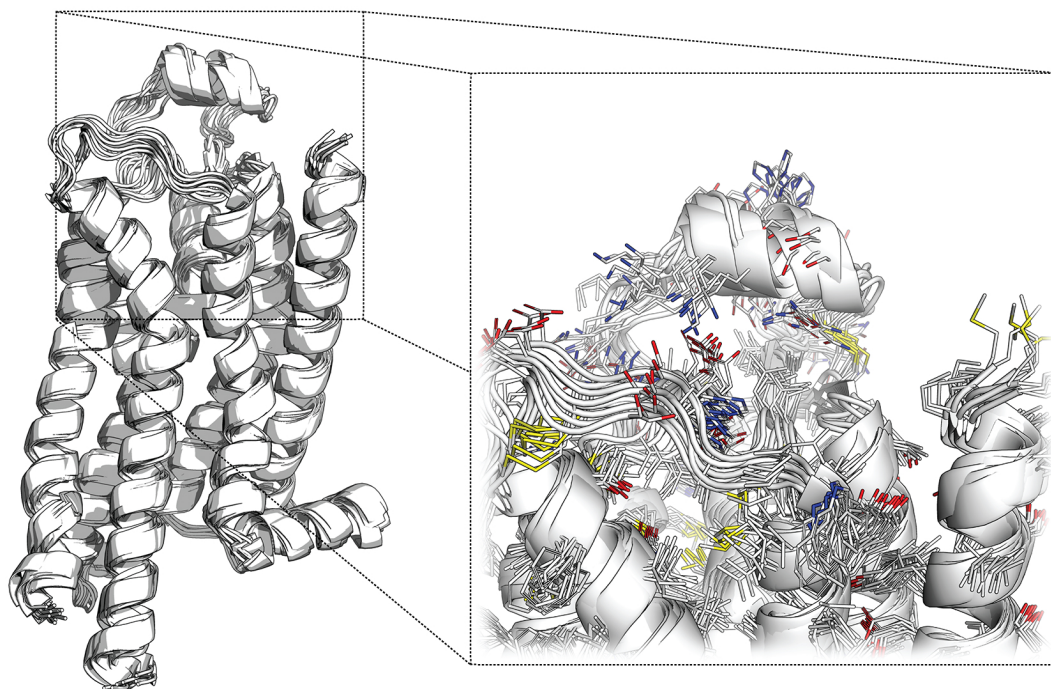
Figure 8: Snapshots of a molecular dynamics simulation of A_{2A} adenosine receptor (5IU4 [137]). The secondary structure and protein backbone can stay stable over a long period of simulation time. In contrast, side chains can adopt various conformation.

3.3   Evaluation of Virtual Screening Performance

The aim of virtual screening experiments is to identify active molecules in databases containing active and inactive molecules. Typically, the performance of a virtual screening protocol is evaluated before the actual experiment by screening a test set comprising of known active and inactive molecules. The resulting hit list is analyzed and hits are classified into four categories (Fig 9), .i.e. true positives (TP), false positives (FP), false negatives (FN) and true negatives (TN) [138]. Several metrics were developed to evaluate and compare virtual screening performance [128, 138, 139].
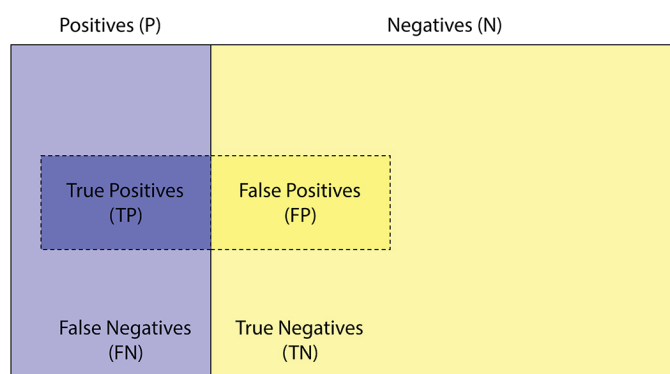
Figure 9: Classification of results from a virtual screening protocol. Blue and yellow areas correspond to the active and inactive molecules of the test, respectively. The dark shaded area represents the molecules predicted to be active by a virtual screening protocol.

The *Sensitivity* (Se, Eq 1) and *Specificity* (Sp, Eq 2) report the fraction of active molecules correctly classified as active and the fraction of inactive molecules correctly classified as inactive, respectively [138].

$$Se = \frac{TP}{P} \tag{1}$$

$$Sp = \frac{TN}{N} \tag{2}$$

The *Yield of Actives* (YA, Eq 3) describes the proportion of actives among all retrieved molecules of the test set. Hence, it can be used to estimate the probability to identify actives in the virtual screening experiment [139].

$$YA = \frac{TP}{TP + FP} \tag{3}$$

The *Enrichment Factor* (EF, Eq 4) allows to determine the enrichment of active molecules in the hit list compared to a random selection [139].

$$EF = \frac{YA}{P/(P+N)} \tag{4}$$

The *receiver operating characteristics* (ROC) plot graphically depicts virtual screening performance by describing *Se* as a function of 1 - *Sp* (Fig 10) [139]. An ideal virtual screening protocol retrieving 100 % of actives and no inactives would yield a ROC curve, which initially rises vertically from the origin to the maximum of *Se*, and than extends to the right horizontally. In contrast, a virtual screening protocol randomly picking active and inactive molecules would be represented by a diagonal spanning the ROC plot from the origin to the top right [128].

Besides a visual interpretation, the ROC plot additionally provides another performance descriptor. The *Area Under the Curve* (AUC, Eq 5) describes the area corresponding to the sum of rectangles formed by Se and 1 - Sp values. Moreover, the AUC calculation can be restricted to a fraction of the ROC plot, e.g. first 10 % of retrieved active and inactive molecules, to better capture the performance of highly scored hits [139].

$$AUC = \sum_i [(Se_{i+1})x(Sp_{i+1} - Sp_i)] \tag{5}$$
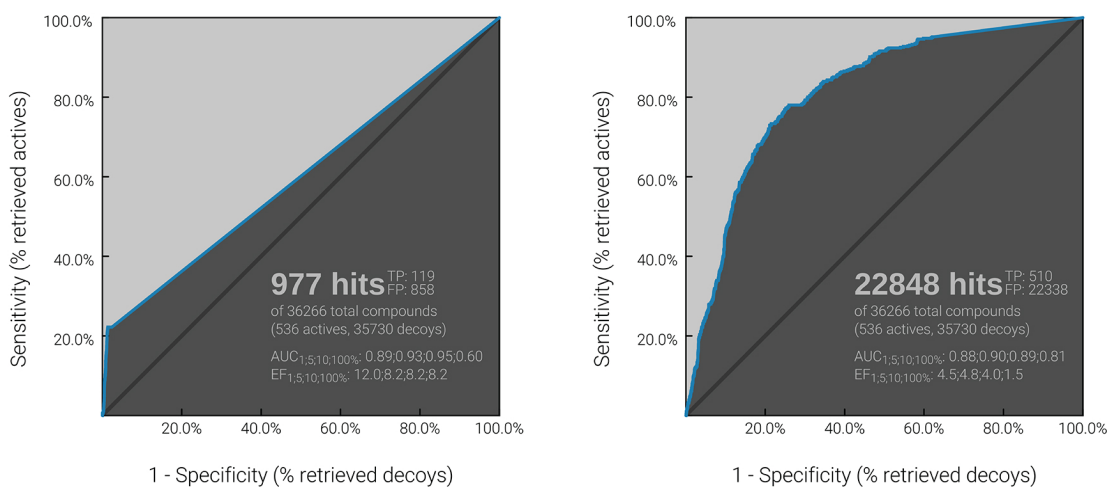
Figure 10: Receiver operating characteristic curves for evaluation of virtual screening performance. TP - true positives, FP - false positives, EF - enrichment factor, AUC - area under the curve.

Results

4.1 Systematic Data Mining Reveals Synergistic H₃R/MCHR1 Ligands

The constantly increasing public availability of bioactivity data together with computationally efficient 2D similarity searches allows for rapid identification of similar molecules with associated activity results. In the following study, bioactivity data was systematically analyzed to identify obesity-relevant target pairs that could be modulated with the same ligand. The most promising target pair comprises H₃R and MCHR1 and was validated *in vitro* by the group of Prof. Dr. Holger Stark at the Heinrich-Heine-University in Düsseldorf, which resulted in the identification of three small molecules with nanomolar affinity for both receptors.

Contribution:

Conceptual design (80 %)
Computational experiments (100 %)
In vitro experiments (0 %)
Visualization (100 %)
Manuscript preparation (80 %)

Systematic Data Mining Reveals Synergistic H3R/MCHR1 Ligands

David Schaller,[†] Stefanie Hagenow,[‡] Gina Alpert,[‡] Alexandra Naß,[†] Robert Schulz,[†] Marcel Bermudez,[†] Holger Stark,[‡] and Gerhard Wolber*[†]

[†]Pharmaceutical and Medicinal Chemistry, Freie Universität Berlin, Königin-Luise-Str. 2+4, 14195 Berlin, Germany
[‡]Pharmaceutical and Medicinal Chemistry, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

Ⓢ Supporting Information

ABSTRACT: In this study, we report a ligand-centric data mining approach that guided the identification of suitable target profiles for treating obesity. The newly developed method is based on identifying target pairs for synergistic positive effects and also encompasses the exclusion of compounds showing a detrimental effect on obesity treatment (off-targets). Ligands with known activity against obesity-relevant targets were compared using fingerprint representations. Similar compounds with activities to different targets were evaluated for the mechanism of action since activation or deactivation of drug targets determines the pharmacological effect. *In vitro* validation of the modeling results revealed that three known modulators of melanin-concentrating hormone receptor 1 (MCHR1) show a previously unknown submicromolar affinity to the histamine H3 receptor (H₃R). This synergistic activity may present a novel therapeutic option against obesity.

KEYWORDS: *Multitarget drugs, fingerprints, histamine H3 receptor, melanin-concentrating hormone receptor 1, obesity*
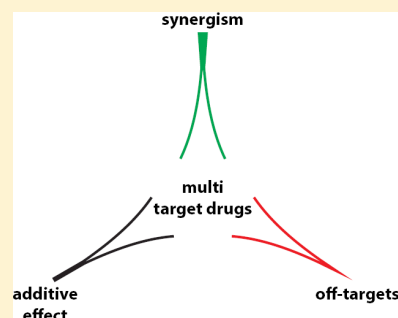
Rational drug design has traditionally focused on the discovery of selective ligands for specific molecular targets. It was assumed that by increasing the selectivity of a ligand for the desired target, undesired side effects arisen from binding to off-targets would be minimized. In recent years, multitarget approaches (often termed "polypharmacology") challenged this dogma proposing that the modulation of multiple targets in the biological network simultaneously may be required to effectively modify a phenotype.[1] Particularly diseases with a complex etiology gained attention for development of multitarget drugs.[2] For instance, several anticancer agents were designed to inhibit certain kinases involved in different aspects of apoptosis and angiogenesis.[3] Also the most effective medications for central nervous system disorders modulate various neurotransmitter levels by targeting several GPCRs or enzymes involved.[4]

Research on databases for ligand activity data indicates that most drugs bind to multiple targets.[5] Furthermore, these drug-target networks are far from being complete since testing each drug against each possible target is economically not favorable. Computational approaches present a suitable option to close this gap and can support the rational multitarget drug design process.[6−9] Analyzing chemical similarities of already known drug-like molecules proved to be particularly successful. Keiser and colleagues were the pioneers in this research field using fingerprint representations of small molecules to predict potential off-targets of approved drugs.[10] Later, Besnard and colleagues calculated Bayesian models for 784 proteins and were able to optimize ligands to a wide array of targets and potential off-targets.[11] Continuously growing public databases

for ligand activity data (e.g., ChEMBL[12]) support these ligand-centric approaches.

In this study, we focused on the first step of rational multitarget drug design, the identification of target pairs that can be modulated by the same ligand. Obesity was chosen as model disease since it is known to bear a complex etiology and single-target medications still lack efficacy and safety.[13] To achieve our goal, we implemented a data mining workflow in KNIME that clusters obesity-relevant targets based on the chemical similarity of ligands from the ChEMBL database.[12,14] Despite its significance for the pharmacological effect, the mechanism of action is still missing for the majority of compounds in public bioactivity databases. For instance, when antagonism of a certain receptor is discussed for obesity treatment, agonism will be ineffective or even induce obesity. Thus, special emphasis was placed on evaluating the mechanism of action of ligand data in terms of activation or deactivation. This strategy led to the identification of several potential target pairs and off-targets that should be considered in obesity treatment. The most promising target pair comprising histamine H3 receptor (H₃R) and melanin-concentrating hormone receptor 1 (MCHR1) could be confirmed *in vitro*.

A literature research yielded 39 obesity-relevant targets with associated activity data stored in the ChEMBL 21 database.[12] These targets can be classified into 25 receptors (24 GPCRs, 1 nuclear receptor), 11 enzymes of the lipid metabolism, and
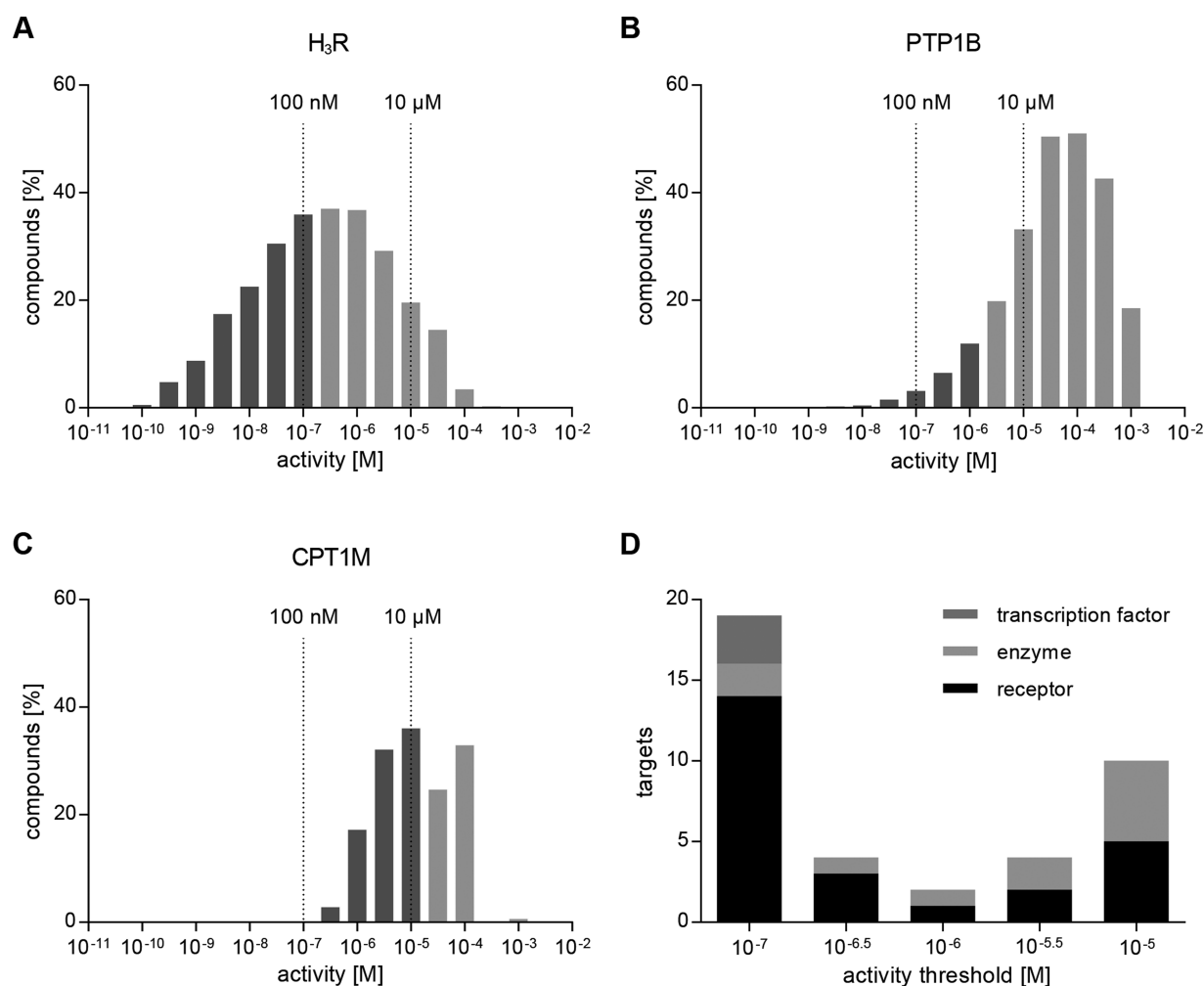
Figure 1. Assignment of activity thresholds for each target separately based on activity data stored in the ChEMBL database. Activity ranges for ligands of (A) histamine H3 receptor, (B) protein-tyrosine phosphatase 1B, and (C) carnitine palmitoyltransferase 1 (muscle isoform). Activity threshold is set three orders of magnitude above the most active compound and limited to a minimum of 100 nM and a maximum of 10 μM. Compounds satisfying the activity threshold are highlighted in dark gray. (D) Distribution of activity thresholds for targets included in this study.

three transcription factors. The multitude of targets discussed in literature underlines the complex etiology of obesity and the necessity to address several targets in the signaling network. A complete list can be found in the Supporting Information (Table S3).

The activity range of ligands in the ChEMBL database can be dramatically different for each target (Figure 1A–C). Thus, a single threshold (e.g., 1 μM) for all targets may not present the most suitable option to extract and focus on the most interesting and active ligands. For instance, a well explored GPCR may require a lower activity threshold than a less well explored protein–protein interaction. Consequently, a protocol has been implemented setting the activity threshold three orders of magnitude above the most active compound. In certain cases, this procedure would result in activity thresholds below 100 nM and subsequently would exclude potentially interesting compounds. Thus, we decided to limit the thresholds to a minimum of 100 nM. Furthermore, a maximum was introduced at 10 μM to exclude poorly active compounds. Targets with a lower activity threshold include several well explored GPCRs, like serotonin receptors, histamine H3 receptor (H_3R), and melanin-concentrating hormone receptor

1 (MCHR1), whereas higher thresholds were commonly assigned to enzymes like carnitine palmitoyltransferase 1 (muscle isoform) and less well explored GPCRs like amylin receptor 1 (Figure 1D). Applying these thresholds to our data set resulted in the selection of 20841 compounds for similarity analysis.

Multitarget action can frequently be observed within a target family since target subtypes bind the same endogenous ligand or substrate and thus share similarities in the binding pocket.[2] Therefore, target subtypes were grouped into target families to allow the identification of more distant relations (structure file activity_data.sdf with assigned target families is provided as Supporting Information).

Subsequently, chemical similarities between compounds of different target families were investigated using Morgan Feature circular fingerprints as implemented in RDKit.[15,16] Compounds were considered similar if they belong to different target families and if the Tanimoto score is 0.7 or higher.

From the initial data set (20841 compounds) only 204 compounds with 233 activities against 19 obesity-relevant targets fulfilled the similarity criteria (Tanimoto score $\geq$ 0.7) to a compound of a different target family.
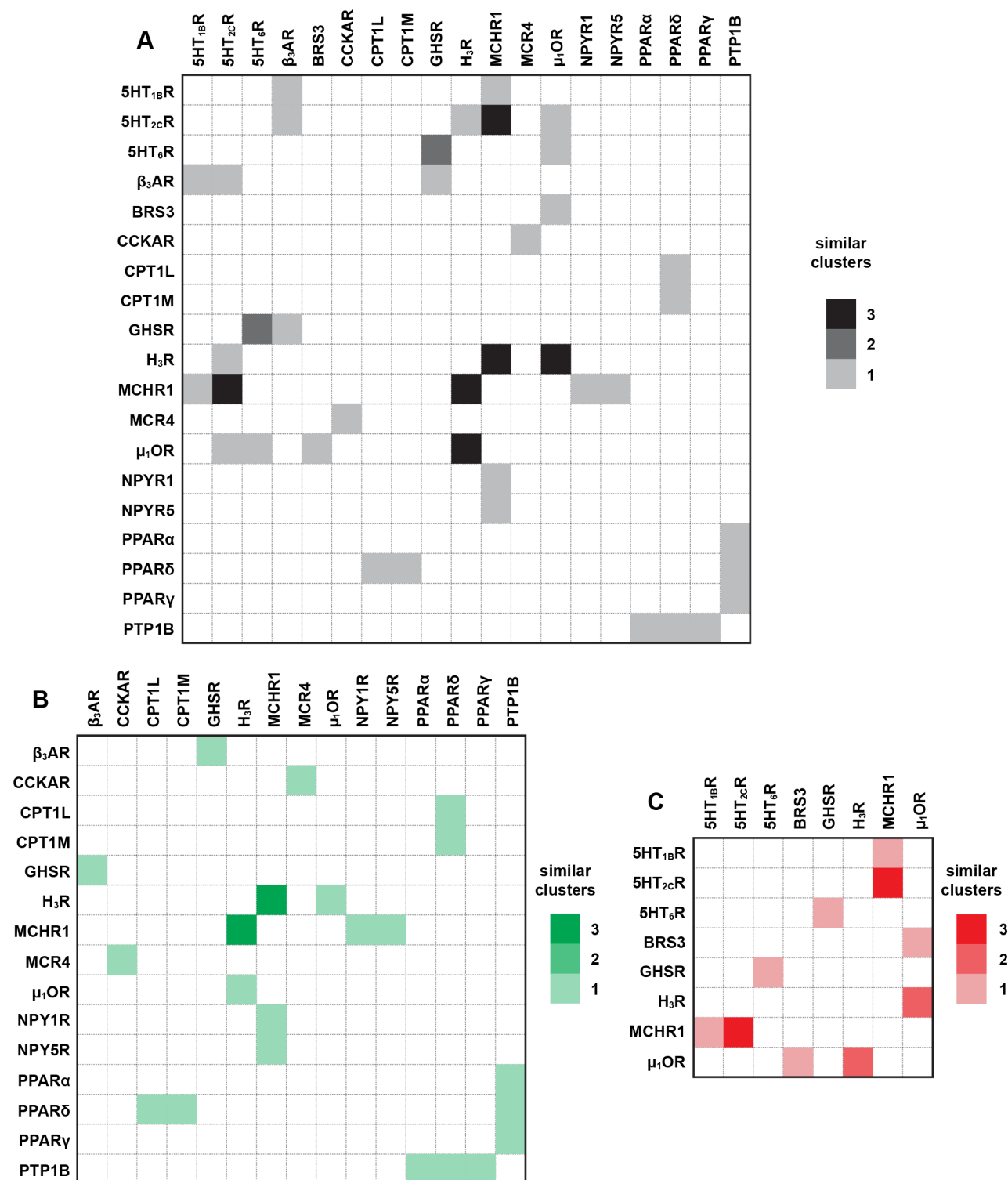
Figure 2. Similarity matrices for obesity-relevant targets based on the chemical similarity of known ligands. Only target pairs are considered that belong to different target families. Ligands were clustered to allow a quality assessment of the target pairs. (A) Similarity matrix without validation of mechanism of action. (B) Similarity matrix with the desired antiobese mechanism of action for both elements of the target pair. (C) Similarity matrix for target pairs, whereas one of the elements of a target pair has a conflictive mechanism of action and thus presents a potential off-target in obesity treatment.

For each target pair, similar compounds were analyzed for diversity by using an in-house implementation of the Taylor–Buti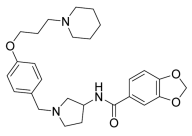na clustering algorithm.[17] This step allows a quality assessment since a higher number of shared similar clusters indicates an increased probability to identify multitarget drugs against this target pair. The identified target pairs are gathered
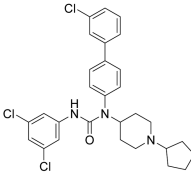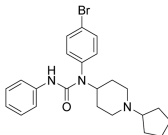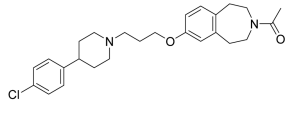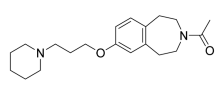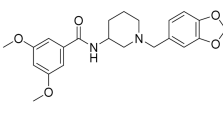
in a similarity matrix, whereas each target pair is rated based on the number of similar clusters (Figure 2A). The target pairs MCHR1/5HT$_{2C}$R, $_{\mu 1}$OR/H$_3$R, and H$_3$R/MCHR1 are rated the best sharing three similar clusters.

Next, the mechanism of action for each cluster was retrieved from literature. This evaluation resulted in the generation of two similarity matrices (Figure 2B,C). One holds information about possible synergistic effects with the desired antiobese mechanism of action for both elements of the target pair (Figure 2B). The target pair comprising H$_3$R and MCHR1 is the only one with more than one similar cluster. The second similarity matrix shows potential off-targets (Figure 2C). For instance, MCHR1 antagonists (desired mechanism of action) show similarities to serotonin receptor 2C (5HT$_{2C}$R) antagonists (conflictive mechanism of action). Noteworthy, several screening campaigns against MCHR1 have reported 5HT$_{2C}$R as off-target.[18] A full list of clusters with associated mechanism of action can be found in the Supporting Information (Table S4).
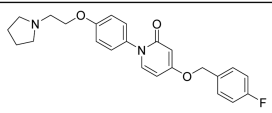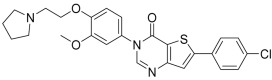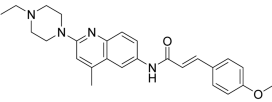
The identified similar clusters for H$_3$R/MCHR1 (Table 1) share a positively charged amine function that is known to be involved in Coulomb interactions with a conserved aspartate for many aminergic GPCRs but also for MCHR1.[18,19]

Table 1. Cluster Pairs Binding to H$_3$R and MCHR1, Respectively

MCHR1	H$_3$R
A CHEMBL433591	CHEMBL1094029
B CHEMBL1914860	CHEMBL210291
C CHEMBL187916	CHEMBL3094128

Considering the high Morgan Feature fingerprint similarity of known ligands, H$_3$R and MCHR1 were chosen for further validation. A shape-based screening campaign using ROCS led to the selection of three known MCHR1 antagonists for a radioligand displacement assay at H$_3$R.[20] All three tested compounds show submicromolar activity against both receptors (Table 2). Compounds **1** and **2** were already described to have antiobesity effects in rodents.[21−24] To our knowledge, compound **3** with the most balanced activity against both receptors (K_i/IC$_{50}$ < 20 nM) has not yet been tested *in vivo*. The ligand efficiency (LE) for compound **3** of 0.34 lies above the limit for drug-like molecules (LE > 0.3) and thus indicates a good starting point for further development.[25] The lip-

Table 2. Activity Table of Known MCHR1 Antagonists

structure	H$_3$R[a] K_i [nM]	MCHR1[b] IC$_{50}$ [nM]
1	516.7	5.6
2	208.0	89.1
3	10.6	16.0

[a]Mean of at least three independent experiments, each performed at least in duplicates in a radioligand displacement assay at H$_3$R. [b]Data for compounds **1**,[24] **2**,[23] and **3**[26] are taken from the literature.

ophilicity-corrected ligand efficiency (LELP) includes lipophilicity for quality assessment as this property has been shown to accompany with promiscuity.[25] The LELP of 12.40 for compound **3** points to potential promiscuity issues. Indeed, closely related compounds of this series show moderate affinity at 5HT$_{2C}$R, emphasizing the consideration of this receptor as off-target.[26]

Only two studies were found describing compounds with a multitarget character against H$_3$R and MCHR1.[27,28] However, the authors did not aim at developing compounds with balanced activity against both receptors. Screening campaigns for selective antagonists of H$_3$R or MCHR1 did not yet result in development of an effective antiobesity treatment. Though, there is evidence for a possible synergistic effect. A recent study revealed that activation of H$_3$R leads to the inhibition of MCH expression.[29] This inhibition could be avoided through administration of a H$_3$R antagonist resulting in expression of MCH. A concurrent expression of the appetite stimulant MCH might explain why the ongoing effort in designing H$_3$R antagonists for obesity treatment did not lead to an effective therapy yet. Although this study focused on sleep and arousal, translating these results into obesity research indicates a promising synergistic effect of dual antagonism of H$_3$R and MCHR1.

In this study, we have successfully applied a ligand-centric data mining approach to identify target pairs that have the potential to drive future multitarget drug research for obesity treatment. The most promising target pair comprising H$_3$R and MCHR1 was validated *in vitro*. Three compounds have been confirmed to hold a multitarget character in the submicromolar activity range. Evaluating the mechanism of action not only allowed the identification of potential target pairs but additionally pointed to several off-targets that should be considered in antiobesity drug development.

■ ASSOCIATED CONTENT

⬤ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsmedchem-lett.7b00118.

Experimental procedures as well as tables of included targets and identified similarities (PDF)
Structures in sdf format (ZIP)

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: gerhard.wolber@fu-berlin.de.
ORCID
David Schaller: 0000-0002-1881-4518
Gerhard Wolber: 0000-0002-5344-0048

Author Contributions
D.S. conducted analysis, designed and performed experiments, and wrote the manuscript. S.H. and G.A. designed and performed experiments. A.N., R.S., M.B., H.S., and G.W. designed experiments. H.S. and G.W. directed the studies. All authors reviewed the manuscript.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ ABBREVIATIONS

$5HT_{1B}R$, serotonin receptor 1B; $5HT_{2C}R$, serotonin receptor 2C; $5HT_6R$, serotonin receptor 6; $\beta 3AR$, beta 3 adrenergic receptor; BRS3, bombesin receptor subtype 3; CCKAR, cholecystokinin A receptor; CPT1L, carnitine O-palmitoyltransferase 1 (liver isoform); CPT1M, carnitine O-palmitoyltransferase 1 (muscle isoform); GHSR, growth hormone secretory receptor; H_3R, histamine H3 receptor; MCR4, melanocortin receptor 4; MCHR1, melanin-concentrating hormone receptor 1; μ_1OR, mu 1 opioid receptor; NPYR1, neuropeptide Y receptor 1; NPYR5, neuropeptide Y receptor 5; PPARα, peroxisome proliferator-activated receptor alpha; PPARδ, peroxisome proliferator-activated receptor delta; PPARγ, peroxisome proliferator-activated receptor gamma; PTP1B, protein-tyrosine phosphatase 1B

■ REFERENCES

(1) Anighoro, A.; Bajorath, J.; Rastelli, G. Polypharmacology: Challenges and Opportunities in Drug Discovery. *J. Med. Chem.* **2014**, *57* (19), 7874−7887.
(2) Peters, J.-U. Polypharmacology - Foe or Friend? *J. Med. Chem.* **2013**, *56* (22), 8955−8971.
(3) Faivre, S.; Demetri, G.; Sargent, W.; Raymond, E. Molecular Basis for Sunitinib Efficacy and Future Clinical Development. *Nat. Rev. Drug Discovery* **2007**, *6* (9), 734−745.
(4) Roth, B. L.; Sheffler, D. J.; Kroeze, W. K. Magic Shotguns versus Magic Bullets: Selectively Non-Selective Drugs for Mood Disorders and Schizophrenia. *Nat. Rev. Drug Discovery* **2004**, *3* (4), 353−359.
(5) Mestres, J.; Gregori-Puigjané, E.; Valverde, S.; Solé, R. V. The Topology of Drug-Target Interaction Networks: Implicit Dependence on Drug Properties and Target Families. *Mol. BioSyst.* **2009**, *5* (9), 1051−1057.
(6) Lavecchia, A.; Cerchia, C. In Silico Methods to Address Polypharmacology: Current Status, Applications and Future Perspectives. *Drug Discovery Today* **2016**, *21* (2), 288−298.
(7) Nikolic, K.; Filipic, S.; Agbaba, D.; Stark, H. Procognitive Properties of Drugs with Single and Multitargeting H 3 Receptor Antagonist Activities. *CNS Neurosci. Ther.* **2014**, *20*, 613−623.
(8) Nikolic, K.; Agbaba, D.; Stark, H. Pharmacophore Modeling, Drug Design and Virtual Screening on Multi-Targeting Procognitive Agents Approaching Histaminergic Pathways. *J. Taiwan Inst. Chem. Eng.* **2015**, *46*, 15−29.
(9) Khanfar, M. A.; Affini, A.; Lutsenko, K.; Nikolic, K.; Butini, S.; Stark, H. Multiple Targeting Approaches on Histamine H3 Receptor Antagonists. *Front. Neurosci.* **2016**, *10* (MAY), 1−17.
(10) Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating Protein Pharmacology by Ligand Chemistry. *Nat. Biotechnol.* **2007**, *25* (2), 197−206.
(11) Besnard, J.; Ruda, G. F.; Setola, V.; Abecassis, K.; Rodriguiz, R. M.; Huang, X.-P.; Norval, S.; Sassano, M. F.; Shin, A. I.; Webster, L. A.; Simeons, F. R. C.; Stojanovski, L.; Prat, A.; Seidah, N. G.; Constam, D. B.; Bickerton, G. R.; Read, K. D.; Wetsel, W. C.; Gilbert, I. H.; Roth, B. L.; Hopkins, A. L. Automated Design of Ligands to Polypharmacological Profiles. *Nature* **2012**, *492* (7428), 215−220.
(12) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42*, D1083.
(13) Saltiel, A. R. New Therapeutic Approaches for the Treatment of Obesity. *Sci. Transl. Med.* **2016**, *8* (323), 323rv2.
(14) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner. *Data Analysis, Machine Learning and Applications* **2008**, 319−326.
(15) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742−754.
(16) RDKit: Open-Source Cheminformatics. http://www.rdkit.org.
(17) Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (4), 747−750.
(18) Högberg, T.; Frimurer, T. M.; Sasmal, P. K. Melanin Concentrating Hormone Receptor 1 (MCHR1) Antagonists - Still a Viable Approach for Obesity Treatment? *Bioorg. Med. Chem. Lett.* **2012**, *22* (19), 6039−6047.
(19) Katritch, V.; Cherezov, V.; Stevens, R. C. Structure-Function of the G Protein−Coupled Receptor Superfamily. *Annu. Rev. Pharmacol. Toxicol.* **2013**, *53* (1), 531−556.
(20) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50* (1), 74−82.
(21) Hertzog, D. L.; Al-Barazanji, K. A.; Bigham, E. C.; Bishop, M. J.; Britt, C. S.; Carlton, D. L.; Cooper, J. P.; Daniels, A. J.; Garrido, D. M.; Goetz, A. S.; Grizzle, M. K.; Guo, Y. C.; Handlon, A. L.; Ignar, D. M.; Morgan, R. O.; Peat, A. J.; Tavares, F. X.; Zhou, H. The Discovery and Optimization of Pyrimidinone-Containing MCH R1 Antagonists. *Bioorg. Med. Chem. Lett.* **2006**, *16* (18), 4723−4727.
(22) Ito, M.; Ishihara, A.; Gomori, A.; Egashira, S.; Matsushita, H.; Mashiko, S.; Ito, J.; Ito, M.; Nakase, K.; Haga, Y.; Iwaasa, H.; Suzuki, T.; Ohtake, N.; Moriya, M.; Sato, N.; MacNeil, D. J.; Takenaga, N.; Tokita, S.; Kanatani, A. Melanin-Concentrating Hormone 1-Receptor Antagonist Suppresses Body Weight Gain Correlated with High Receptor Occupancy Levels in Diet-Induced Obesity Mice. *Eur. J. Pharmacol.* **2009**, *624* (1−3), 77−83.
(23) Oyarzabal, J.; Howe, T.; Alcazar, J.; Andrés, J. I.; Alvarez, R. M.; Dautzenberg, F.; Iturrino, L.; Martínez, S.; Van der Linden, I. Novel Approach for Chemotype Hopping Based on Annotated Databases of Chemically Feasible Fragments and a Prospective Case Study: New Melanin Concentrating Hormone Antagonists. *J. Med. Chem.* **2009**, *52* (7), 2076−2089.
(24) Haga, Y.; Mizutani, S.; Naya, A.; Kishino, H.; Iwaasa, H.; Ito, M.; Ito, J.; Moriya, M.; Sato, N.; Takenaga, N.; Ishihara, A.; Tokita, S.; Kanatani, A.; Ohtake, N. Discovery of Novel Phenylpyridone

Derivatives as Potent and Selective MCH1R Antagonists. *Bioorg. Med. Chem.* **2011**, *19* (2), 883−893.

(25) Hopkins, A. L.; Keserü, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H. The Role of Ligand Efficiency Metrics in Drug Discovery. *Nat. Rev. Drug Discovery* **2014**, *13*, 105.

(26) Ulven, T.; Frimurer, T. M.; Receveur, J.-M.; Little, P. B.; Rist, O.; Nørregaard, P. K.; Högberg, T. 6-Acylamino-2-Aminoquinolines as Potent Melanin-Concentrating Hormone 1 Receptor Antagonists. Identification, Structure-Activity Relationship, and Investigation of Binding Mode. *J. Med. Chem.* **2005**, *48* (18), 5684−5697.

(27) Cirauqui, N.; Schrey, A. K.; Galiano, S.; Ceras, J.; Pérez-Silanes, S.; Aldana, I.; Monge, A.; Kühne, R. Building a MCHR1 Homology Model Provides Insight into the Receptor-Antagonist Contacts That Are Important for the Development of New Anti-Obesity Agents. *Bioorg. Med. Chem.* **2010**, *18* (21), 7365−7379.

(28) Johansson, A.; Löfberg, C.; Antonsson, M.; Von Unge, S.; Hayes, M. A.; Judkins, R.; Ploj, K.; Benthem, L.; Lindén, D.; Brodin, P.; Wennerberg, M.; Fredenwall, M.; Li, L.; Persson, J.; Bergman, R.; Pettersen, A.; Gennemark, P.; Hogner, A. Discovery of (3-(4-(2-Oxa-6-azaspiro[3.3]heptan-6-Ylmethyl)phenoxy)azetidin-1-yl)(5-(4-Methoxyphenyl)-1,3,4-Oxadiazol-2-Yl)methanone (AZD1979), a Melanin Concentrating Hormone Receptor 1 (MCHr1) Antagonist with Favorable Physicochemical Properties. *J. Med. Chem.* **2016**, *59* (6), 2497−2511.

(29) Parks, G. S.; Olivas, N. D.; Ikrar, T.; Sanathara, N. M.; Wang, L.; Wang, Z.; Civelli, O.; Xu, X. Histamine Inhibits the Melanin-Concentrating Hormone System: Implications for Sleep and Arousal. *J. Physiol.* **2014**, *592* (Pt 10), 2183−2196.

Supporting Information

Systematic data mining reveals synergistic H3R/MCHR1 ligands

David Schaller, Stefanie Hagenow, Gina Alpert, Alexandra Naß, Robert Schulz, Marcel Bermudez, Holger Stark and Gerhard Wolber[*]

Table of contents

Experimental Procedures

Data mining

The PubMed database was searched for reviews that contain the keywords "obesity" and "treatment" in the title or abstract.[1] Discussed targets were further reviewed and checked for available activity data in the ChEMBL 21 database.[2] This procedure yielded 39 obesity-relevant targets (Tab. S3).

The following workflow was conducted in KNIME if not specified else.[3] The activity data of studied targets was extracted from the ChEMBL 21 database.[2] Several criteria were applied to exclude ambiguous data. Compounds were filtered for confidence score ($\geq$7), organism (homo sapiens), activity type (K_i, K_D, IC_{50} or EC_{50}), standard units (nM) and operator (=). Additionally, a molecular weight cutoff was set to 700 Da. In total 36626 compounds with 56740 activity data points were included in this study. If multiple data points were available for one compound against the same target, binding data (K_i, K_D) was preferred over functional data (IC_{50}, EC_{50}) and more recent published data was preferred over older data. This procedure condensed the activity data to 41545 activities.

Activity thresholds were set for each target separately based on the available data in the ChEMBL 21 database.[2] The threshold was set three orders of magnitude above the most active compound. However, if a threshold would fall below 100 nM or above 10 μM, this threshold is set to 100 nM or 10 μM respectively (Fig. 2). 20841 compounds with 22018 activities against 38 targets remained for further analysis. The pancreatic lipase was excluded because the compounds did not match the filtering criteria.

38 target subtypes were grouped into 26 target families to focus on more distant relations. For instance, $5HT_{2C}R$, $5HT_{1B}R$ and $5HT_6R$ are part of 5HTR.

Compounds were protonated and fragments removed by using the database wash application in MOE (structure file activity_data.sdf is provided as supporting information).[4] MorganFeat fingerprints (diameter = 4) were generated using the RDKit.[5,6] Compounds were considered similar if the Tanimoto score was at least 0.7 and if they belong to different target families. This procedure resulted in the retrieval of 204 compounds with 233 activities.

To allow quality assessment, similar molecules for each target pair were clustered using an in-house implementation of the Taylor-Butina algorithm with a Tanimoto cutoff at 0.5.[7] First, the number of neighbors (similar molecules with a Tanimoto score of at least 0.5) is calculated for each molecule. Next, the molecule with the most neighbors and all its neighbors are used to define the first cluster. Then, the prior steps are run a second time without the molecules of the first cluster to define the second cluster. These steps are repeated until all molecules are assigned to a cluster. Cluster pairs are defined in each cluster separately by identifying those molecules that show activity against different targets and are the most similar for the investigated cluster. Finally, each element of the cluster pairs was evaluated for their mechanism of action in terms of activation or deactivation (e.g. agonist or antagonist, Tab. S4). Furthermore, target pairs were investigated for potential synergistic effects.
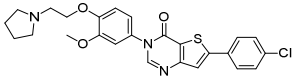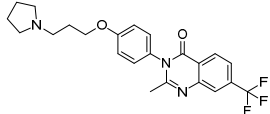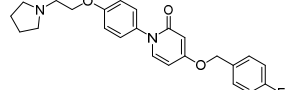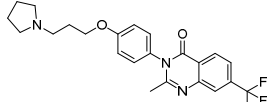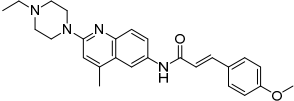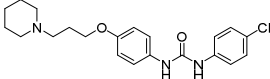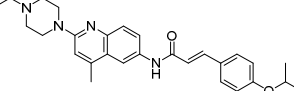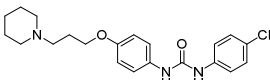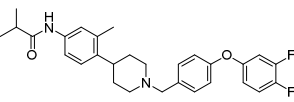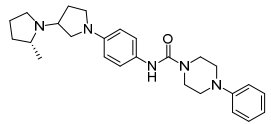
Virtual Screening

Based on our results MCHR1 and H_3R presented the most promising target pair and were chosen for further investigation. Scientific literature has been screened for publications to compile manually curated databases of known H_3R and MCHR1 antagonists. Structures were downloaded from ChEMBL or if not ava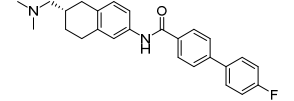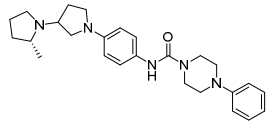ilable using the Sdf-Export tool from Scifinder.[2,8] Six known MCHR1 antagonists (K_i/IC_{50} <= 100 nM) were found in the ZINC database to be purchasable from different vendors (structure file ZINC_MCHR1.sdf is provided as supporting information).[9] These compounds were used as query in a shape-based screening campaign against 342 known

H$_3$R antagonists (K$_i$/IC$_{50}$ <= 1 nM, structure file H3R_1nM.sdf is provided as supporting information). First, MCHR1 and H$_3$R antagonists were protonated and energy minimized with the MMFF94 forcefield in MOE.[4,10] Next, conformations of the H$_3$R antagonists were generated using OMEGA with default settings.[11] Finally, each purchasable MCHR1 antagonist was screened against H$_3$R antagonist conformations using ROCS with default settings.[12] Results were analyzed using the TanimotoCombo score as implemented in ROCS (Tab. S1).
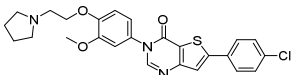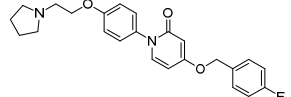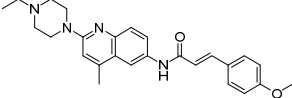
Compound **1 − 5** possess a TanimotoCombo score higher than 1.0 and were considered for *in-vitro* validation. Compound **3** and **4** are nearly identical. Thus, compound **4** with the lower score was excluded. Compound **5** turned out to be not in stock and was not purchased. Compound **6** had a score below 1.0 and hence was not considered for *in-vitro* validation.

Table S1: Purchasable MCHR1 antagonists (K$_i$/IC$_{50}$ <= 100 nM) and the most similar H$_3$R antagonist (K$_i$/IC$_{50}$ <= 1 nM) according to the TanimotoCombo score.

	MCHR1	TanimotoCombo	H$_3$R
1		1.441	
2		1.295	
3		1.165	
4		1.099	
5		1.386	
6		0.979	

Ordered compounds were analyzed with LC-MS and possess a purity of at least 95 % (Tab. S2).

Table S2. Purity and activity of tested compounds analyzed with LC-MS and radioligand depletion experiment, respectively.

	structure	purity	MW [g/mol]	m/z [M+H]$^+$	pK$_i$ ± SEM
1		> 95 %	408.2	409.0	6.34 ± 0.14
2		> 95 %	481.1	481.8	6.70 ± 0.07
3		> 95 %	430.2	431.0	7.97 ± 0.64

Ligand efficiency calculations

Ligand efficiency (LE) and lipophilicity-corrected ligand efficiency (LELP) were calculated as published previously.[13] The clogP of 4.217 for compound **3** used for LELP was calculated using MOE.[4]

Histamine H$_3$ receptor *in-vitro* assay

For preparation of crude hH$_3$R membrane extracts HEK-293 cells stably expressing the hH$_3$R were cultivated, harvested and processed as described previously.[14]

For the radioligand depletion experiments cell membranes were thawed and homogenized by sonication in ice-cold binding buffer (12.5mM MgCl2,1mM EDTA and 75mM Tris/HCl, pH 7.4). Crude membrane extracts (20 µg/well; final volume of 0.2 ml) were incubated (90 min; room temperature) with various concentrations of test ligands (between 0.01 nM and 100 µM) and [^{3}H]-N-alpha-methylhistamine (2 nM final concentration; 78.3 Ci/mmol). Nonspecific binding was obtained by using pitolisant (10 µM final concentration). Membrane extracts were separated from unbound components by filtration through GF/B filters pre-treated with 0.3% (m/v) polyethyleneimine using an Inotech cell harvester. Liquid scintillation counting was used

for measuring bound radioligand. Data analysis were performed with GraphPad Prism 6 using non-linear regression. The Ki values for each experiment were obtained by using an incorporated equation of GraphPad Prism according to Cheng-Prusoff. Statistical analysis was conducted on pK_i values. Mean values were calculated from at least three independent experiments, each performed at least in duplicates (Tab. S1). Confidence intervals (95%) were calculated and converted to nanomolar concentrations.

Obesity-relevant targets

Table S3. Obesity-relevant targets included in this study categorized by the desired mechanism of action to induce an anti-obese effect.

Agonists/activators

Target	Full name	CHEMBLID	Target family
$5HT_{1B}R^{15}$	Serotonin 1B receptor	1898	5HTR
$5HT_{2C}R^{15}$	Serotonin 2C receptor	225	5HTR
AMY1[16]	Amylin receptor 1	2111189	AMY
AMY3[16]	Amylin receptor 3	2111190	AMY
β_3AR	Beta 3 adrenergic receptor	246	β3AR
BRS3[17]	Bombesin receptor subtype 3	4080	BRS3
CCKAR[18]	Cholecystokinin A receptor	1901	CCKAR
GLP1R[18]	Glucagon-like peptide 1 receptor	1784	GLP1R
MCR3[19]	Melanocortin receptor 3	4644	MCR
MCR4[19]	Melanocortin receptor 4	259	MCR
NPYR2[20]	Neuropeptide Y receptor 2	4018	NPYR
NPYR4[20]	Neuropeptide Y receptor 4	4877	NPYR
OXR1[21]	Orexin receptor 1	5113	OXR
OXR2[21]	Orexin receptor 2	4792	OXR
PPARα[22]	Peroxisome proliferator-activated receptor alpha	239	PPAR
PPARδ[22]	Peroxisome proliferator-activated receptor delta	3979	PPAR
PPARγ[22]	Peroxisome proliferator-activated receptor gamma	235	PPAR
SIRT1[23]	Sirtuin 1	4506	SIRT1
THRβ[24]	Thyroid hormone receptor beta	1947	THRβ

Antagonists/inverse agonists/inhibitors

Target	Full name	CHEMBLID	Target family
11βHD1[25]	11-beta hydroxysteroid dehydrogenase 1	4235	11βHD1
$5HT_6R^{26}$	Serotonin 6 receptor	3371	5HTR
ACC1[27]	Acetyl-CoA carboxylase 1	3351	ACC
ACC2[27]	Acetyl-CoA carboxylase 2	4829	ACC
SCD1[28]	Stearoyl-CoA desaturase 1	5555	SCD1
CB1[29]	Cannabinoid receptor 1	218	CB1
CPT1L[30]	Carnitine O-palmitoyltransferase 1, liver isoform	1293194	CPT1

CPT1M[30]	Carnitine O-palmitoyltransferase 1, muscle isoform	2216739	CPT1
CRHR2[31]	Corticotropin releasing hormone receptor 2	4096	CRHR2
DGAT1[32]	Diacylglycerol O-acyltransferase	6009	DGAT1
FAS[33]	Fatty acid synthase	4158	FAS
GALR1[34]	Galanin receptor 1	4894	GAL1R
GHSR[18]	Growth hormone secretory receptor	4616	GHSR
H_3R[35]	Histamine H3 receptor	264	H3R
MCHR1[36]	Melanin-concentrating hormone receptor 1	344	MCHR1
μ_1OR[37]	Mu 1 opioid receptor	233	MOR
NPYR1[20]	Neuropeptide Y receptor 1	4777	NPYR
NPYR5[20]	Neuropeptide Y receptor 5	4561	NPYR
PLIP[38]	Pancreatic lipase	1812	PLIP
PTP1B[39]	Protein-tyrosine phosphatase 1B	335	PTP1B

Cluster pairs with mode of action

Table S4. Similar cluster pairs with mechanism of action (MOA) and activity to different obesity-relevant targets. n/a – mode of action not available in literature.

target	CHEMBL ID	MOA	tanimoto	MOA	CHEMBL ID	target
$5HT_{1B}R$	3126382	n/a	0.82	agonist	1814275	$\beta_3 AR$
$5HT_{1B}R$	194837	antagonist	1.00	antagonist	194837	MCHR1
$5HT_{2C}R$	723	n/a	1.00	antagonist	723	$\beta_3 AR$
$5HT_{2C}R$	127307	agonist	1.00	n/a	127307	$H_3 R$
$5HT_{2C}R$	1818901	antagonist	1.00	antagonist	1818800	MCHR1
$5HT_{2C}R$	216280	antagonist	1.00	antagonist	216280	MCHR1
$5HT_{2C}R$	383800	antagonist	0.71	antagonist	215508	MCHR1
$5HT_{2C}R$	482496	agonist	1.00	n/a	482496	$\mu_1 OR$
$5HT_6 R$	431298	antagonist	0.83	agonist	2364345	GHSR1a
$5HT_6 R$	1079311	n/a	1.00	agonist	1079311	GHSR1a
$5HT_6 R$	482496	agonist	1.00	n/a	482496	$\mu_1 OR$
$\beta_3 AR$	1814275	agonist	0.82	n/a	3126382	$5HT_{1B}R$
$\beta_3 AR$	723	antagonist	1.00	n/a	723	$5HT_{2C}R$
$\beta_3 AR$	12998	agonist	0.87	antagonist	1077617	GHSR1a
BRS3	3144501	agonist	0.81	agonist	2178733	$\mu_1 OR$
CCKAR	327815	agonist	0.77	agonist	591041	MCR4
CPT1L	3431630	inhibitor	0.85	agonist	522575	PPARδ
CPT1M	3431628	inhibitor	0.76	agonist	496116	PPARδ
GHSR	2364345	agonist	0.83	antagonist	431298	$5HT_6 R$
GHSR	1079311	agonist	1.00	n/a	1079311	$5HT_6 R$
GHSR	1077617	antagonist	0.87	agonist	12998	$\beta_3 AR$
$H_3 R$	127307	n/a	1.00	agonist	127307	$5HT_{2C}R$
$H_3 R$	1094029	antagonist	0.88	antagonist	433591	MCHR1
$H_3 R$	210291	antagonist	0.73	antagonist	1914860	MCHR1
$H_3 R$	3094128	antagonist	0.76	antagonist	187916	MCHR1
$H_3 R$	3092839	antagonist	0.83	antagonist	237294	$\mu_1 OR$
$H_3 R$	441705	antagonist	0.75	agonist	101454	$\mu_1 OR$
$H_3 R$	1627	antagonist	1.00	agonist	1627	$\mu_1 OR$

MCR4	591041	agonist	0.77	agonist	327815	CCKAR
MCHR1	194837	antagonist	1.00	antagonist	194837	$5HT_{1B}R$
MCHR1	1818800	antagonist	1.00	antagonist	1818901	$5HT_{2C}R$
MCHR1	216280	antagonist	1.00	antagonist	216280	$5HT_{2C}R$
MCHR1	215508	antagonist	0.71	antagonist	383800	$5HT_{2C}R$
MCHR1	433591	antagonist	0.88	antagonist	1094029	H_3R
MCHR1	1914860	antagonist	0.73	antagonist	210291	H_3R
MCHR1	187916	antagonist	0.76	antagonist	3094128	H_3R
MCHR1	217171	antagonist	0.72	antagonist	41457	NPYR1
MCHR1	180003	antagonist	1.00	antagonist	193771	NPYR5
μ_1OR	482496	n/a	1.00	agonist	482496	$5HT_{2C}R$
μ_1OR	482496	n/a	1.00	agonist	482496	$5HT_6R$
μ_1OR	2178733	agonist	0.81	agonist	3144501	BRS3
μ_1OR	237294	antagonist	0.83	antagonist	3092839	H_3R
μ_1OR	101454	agonist	0.75	antagonist	441705	H_3R
μ_1OR	1627	agonist	1.00	antagonist	1627	H_3R
NPYR1	41457	antagonist	0.72	antagonist	217171	MCHR1
NPYR5	193771	antagonist	1.00	antagonist	180003	MCHR1
PPARα	1935608	agonist	1.00	inhibitor	1935608	PTP1B
PPARδ	522575	agonist	0.85	inhibitor	3431630	CPT1L
PPARδ	496116	agonist	0.76	inhibitor	3431628	CPT1M
PPARδ	37495	agonist	0.74	inhibitor	282113	PTP1B
PPARγ	1933093	agonist	1.00	inhibitor	1933093	PTP1B
PTP1B	1935608	inhibitor	1.00	agonist	1935608	PPARα
PTP1B	282113	inhibitor	0.74	agonist	37495	PPARδ
PTP1B	1933093	inhibitor	1.00	agonist	1933093	PPARγ

References

(1)   PubMed: US National Library of Medicine, https://www.ncbi.nlm.nih.gov/pubmed/.

(2)   Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42*.

(3)   Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner; 2008; pp 319–326.

(4)   Chemical Computing Group Inc. Molecular Operating Environment (MOE). Montreal, QC, Canada 2015, p 1010 Sherbooke St. West, Suite #910.

(5)   Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754.

(6)   RDKit: Open-Source Cheminformatics; Http://www.rdkit.org.

(7)   Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (4), 747–750.

(8)   Scifinder - A CAS Solution; Http://scifinder.cas.org/scifinder.

(9)   Irwin, J. J.; Shoichet, B. K. ZINC--a Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model. 45* (1), 177–182.

(10) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17* (5–6), 490–519.

(11)   OMEGA 2.5.1.4: OpenEye Scientific Software, Santa Fe, NM; Http://www.eyesopen.com.

(12)   ROCS 3.2.0.4: OpenEye Scientific Software, Santa Fe, NM; Http://www.eyesopen.com.

(13)   Hopkins, A. L.; Keserü, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H. The Role of

          Ligand Efficiency Metrics in Drug Discovery. *Nat. Publ. Gr.* **2014**, *13*.

(14)   Kottke, T.; Sander, K.; Weizel, L.; Schneider, E. H.; Seifert, R.; Stark, H. Receptor-

          Specific Functional Efficacies of Alkyl Imidazoles as Dual Histamine H3/H4 Receptor

          Ligands. *Eur. J. Pharmacol.* **2011**, *654* (3), 200–208.

(15)   Heisler, L. K.; Jobst, E. E.; Sutton, G. M.; Zhou, L.; Borok, E.; Thornton-Jones, Z.; Liu, H.

          Y.; Zigman, J. M.; Balthasar, N.; Kishi, T.; Lee, C. E.; Aschkenasi, C. J.; Zhang, C.-Y.;

          Yu, J.; Boss, O.; Mountjoy, K. G.; Clifton, P. G.; Lowell, B. B.; Friedman, J. M.; Horvath,

          T.; Butler, A. a; Elmquist, J. K.; Cowley, M. a. Serotonin Reciprocally Regulates

          Melanocortin Neurons to Modulate Food Intake. *Neuron* **2006**, *51* (2), 239–249.

(16)   Roth, J. D. Amylin and the Regulation of Appetite and Adiposity: Recent Advances in

          Receptor Signaling, Neurobiology and Pharmacology. *Curr. Opin. Endocrinol. Diabetes.

          Obes.* **2013**, *20* (1), 8–13.

(17)   Ramos-Álvarez, I.; Martín-Duce, A.; Moreno-Villegas, Z.; Sanz, R.; Aparicio, C.; Portal-

          Núñez, S.; Mantey, S. A.; Jensen, R. T.; González, N. Bombesin Receptor Subtype-3

          (BRS-3), a Novel Candidate as Therapeutic Molecular Target in Obesity and Diabetes.

          Mol. Cell. Endocrinol. **2013**, *367*, 109–115.

(18)   Perry, B.; Wang, Y. Appetite Regulation and Weight Control: The Role of Gut Hormones.

          Nutr. Diabetes **2012**, *2* (1), e26.

(19)   Warne, J. P.; Xu, A. W. Metabolic Transceivers: In Tune with the Central Melanocortin

System. *Trends Endocrinol. Metab.* **2013**, *24* (2), 68–75.

(20)   Ishihara PhD, A.; Moriya PhD, M.; MacNeil PhD, D. J.; Fukami PhD, T.; Kanatani PhD,

A. Neuropeptide Y Receptors as Targets of Obesity Treatment. *Expert Opin. Ther. Pat.*

2006, *16* (12), 1701–1712.

(21)   Xu, T.-R.; Yang, Y.; Ward, R.; Gao, L.; Liu, Y. Orexin Receptors: Multi-Functional

Therapeutic Targets for Sleeping Disorders, Eating Disorders, Drug Addiction, Cancers

and Other Physiological Disorders. *Cell. Signal.* **2013**, *25* (12), 2413–2423.

(22)   Grygiel-Górniak, B. Peroxisome Proliferator-Activated Receptors and Their Ligands:

Nutritional and Clinical Implications--a Review. *Nutr. J.* **2014**, *13*, 17.

(23)   Boutant, M.; Joffraud, M.; Kulkarni, S. S.; García-Casarrubios, E.; García-Roves, P. M.;

Ratajczak, J.; Fernández-Marcos, P. J.; Valverde, A. M.; Serrano, M.; Cantó, C. SIRT1

Enhances Glucose Tolerance by Potentiating Brown Adipose Tissue Function. *Mol.*

Metab. **2015**, *4* (2), 118–131.

(24)   Grover, G. J.; Mellström, K.; Malm, J. Therapeutic Potential for Thyroid Hormone

Receptor-Beta Selective Agonists for Treating Obesity, Hyperlipidemia and Diabetes.

Curr. Vasc. Pharmacol. **2007**, *5* (2), 141–154.

(25)   Chapman, K.; Holmes, M.; Seckl, J. 11B-Hydroxysteroid Dehydrogenases: Intracellular

Gate-Keepers of Tissue Glucocorticoid Action. *Physiol. Rev.* **2013**, *93*, 1139–1206.

(26)   Heal, D. J.; Smith, S. L.; Fisas, A.; Codony, X.; Buschmann, H. Selective 5-HT6 Receptor

Ligands: Progress in the Development of a Novel Pharmacological Approach to the

Treatment of Obesity and Related Metabolic Disorders. *Pharmacol. Ther.* **2008**, *117* (2),

207–231.

(27)   Strable, M. S.; Ntambi, J. M. Genetic Control of de Novo Lipogenesis: Role in Diet-Induced Obesity. *Crit. Rev. Biochem. Mol. Biol.* **2010**, *45* (3), 199–214.

(28)   Sampath, H.; Ntambi, J. M. Role of Stearoyl-CoA Desaturase-1 in Skin Integrity and Whole Body Energy Balance. *Journal of Biological Chemistry*. 2014, pp 2482–2488.

(29)   Watkins, B. A.; Kim, J. The Endocannabinoid System: Directing Eating Behavior and Macronutrient Metabolism. *Front. Psychol.* **2014**, *5*, 1506.

(30)   Bruce, C. R.; Hoy, A. J.; Turner, N.; Watt, M. J.; Allen, T. L.; Carpenter, K.; Cooney, G. J.; Febbraio, M. a; Kraegen, E. W. Overexpression of Carnitine Palmitoyltransferase-1 in Skeletal Muscle Is Sufficient to Enhance Fatty Acid Oxidation and Improve High-Fat Diet-Induced Insulin Resistance. *Diabetes* **2009**, *58* (3), 550–558.

(31)   Mastorakos, G.; Zapanti, E. The Hypothalamic-Pituitary-Adrenal Axis in the Neuroendocrine Regulation of Food Intake and Obesity: The Role of Corticotropin Releasing Hormone. *Nutr. Neurosci. 7* (5–6), 271–280.

(32)   Yen, C.-L. E.; Stone, S. J.; Koliwad, S.; Harris, C.; Farese, R. V. Thematic Review Series: Glycerolipids. DGAT Enzymes and Triacylglycerol Biosynthesis. *J. Lipid Res.* **2008**, *49* (11), 2283–2301.

(33)   Lodhi, I. J.; Yin, L.; Jensen-Urstad, A. P. L.; Funai, K.; Coleman, T.; Baird, J. H.; El Ramahi, M. K.; Razani, B.; Song, H.; Fu-Hsu, F.; Turk, J.; Semenkovich, C. F. Inhibiting Adipose Tissue Lipogenesis Reprograms Thermogenesis and PPARγ Activation to Decrease Diet-Induced Obesity. *Cell Metab.* **2012**, *16* (2), 189–201.

(34)   Fang, P.; Yu, M.; Guo, L.; Bo, P.; Zhang, Z.; Shi, M. Galanin and Its Receptors: A Novel Strategy for Appetite Control and Obesity Therapy. *Peptides*. 2012, pp 331–339.

(35)   Masaki, T.; Yoshimatsu, H. Therapeutic Approach of Histamine H3 Receptors in Obesity. *Recent Pat. CNS Drug Discov.* **2007**, *2* (3), 238–240.

(36)   Borowsky, B.; Durkin, M. M.; Ogozalek, K.; Marzabadi, M. R.; DeLeon, J.; Lagu, B.; Heurich, R.; Lichtblau, H.; Shaposhnik, Z.; Daniewska, I.; Blackburn, T. P.; Branchek, T. a; Gerald, C.; Vaysse, P. J.; Forray, C. Antidepressant, Anxiolytic and Anorectic Effects of a Melanin-Concentrating Hormone-1 Receptor Antagonist. *Nat. Med.* **2002**, *8* (8), 825–830.

(37)   Karlsson, H. K.; Tuominen, L.; Tuulari, J. J.; Hirvonen, J.; Parkkola, R.; Helin, S.; Salminen, P.; Nuutila, P.; Nummenmaa, L. Obesity Is Associated with Decreased -Opioid But Unaltered Dopamine D2 Receptor Availability in the Brain. *J. Neurosci.* **2015**, *35*, 3959–3965.

(38)   Hvizdos, K. M.; Markham, A. Orlistat: A Review of Its Use in the Management of Obesity. *Drugs* **1999**, *58* (4), 743–760.

(39)   Allison, M. B.; Myers, M. G. 20 YEARS OF LEPTIN: Connecting Leptin Signaling to Biological Function. *J. Endocrinol.* **2014**, *223* (1), T25–T35.

4.2 Ligand-Guided Homology Modeling Drives Identification of Novel Histamine H₃ Receptor Ligands

The systematic data mining approach revealed several potential target pairs for anti-obesity drug development (see section 4.1). The most promising target pair, consisting of H₃R and MCHR1, was validated *in vitro*. The testing of known MCHR1 antagonists for H₃R binding resulted in the identification of three dual modulators with nanomolar binding affinity. However, binding mode investigations, which could facilitate rational lead optimization are hampered by the absence of experimentally resolved atomistic models. Hence, we aimed at for predicting the atomistic structures of both receptors, which could assist the challenging optimization against multiple targets. In the following study, the generation of H₃R atomistic models was driven by a ligand-guided homology modeling approach with subsequent 3D pharmacophore-based virtual screening. Identified compounds were validated *in vitro* for H₃R binding by the group of Prof. Dr. Holger Stark at the Heinrich-Heine-University in Düsseldorf, which resulted in the identification of two novel compounds with nanomolar affinity.

Contribution:

Conceptual design (80 %)
Computational experiments (100 %)
In vitro experiments (0 %)
Visualization (100 %)
Manuscript preparation (80 %)

RESEARCH ARTICLE

Ligand-guided homology modeling drives identification of novel histamine H3 receptor ligands

David Schaller[1], Stefanie Hagenow[2], Holger Stark[2], Gerhard Wolber[1]*

1 Molecular Design Lab, Pharmaceutical and Medicinal Chemistry, Institute of Pharmacy, Freie Universität Berlin, Berlin, Germany, 2 Institute of Pharmaceutical and Medicinal Chemistry, Department of Pharmacy, Faculty of Mathematics and Natural Sciences, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Nordrhein-Westfalen, Germany

* gerhard.wolber@fu-berlin.de
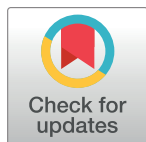
Abstract

In this study, we report a ligand-guided homology modeling approach allowing the analysis of relevant binding site residue conformations and the identification of two novel histamine H_3 receptor ligands with binding affinity in the nanomolar range. The newly developed method is based on exploiting an essential charge interaction characteristic for aminergic G-protein coupled receptors for ranking 3D receptor models appropriate for the discovery of novel compounds through virtual screening.

Introduction

Virtual screening campaigns are typically classified into ligand-based approaches exploiting the similarity of molecules to already known active ligands, and structure-based approaches, where virtual screening models describe three-dimensional chemical interactions between molecules and the target structure [1]. A literature survey revealed that structure-based approaches are on average less successful in identifying highly active hits than ligand-based approaches [2]. However, if active lead compounds are identified, structure-based approaches hold the information for a subsequent rational optimization of interactions between ligand and target structure.

Although the amount of publicly available data for ligand-protein complexes is constantly increasing, structural data is not always available. In this situation researchers often rely on homology modeling, a method for generating the protein structure of interest based on closely related proteins with resolved crystal structures [3]. Including ligand information can aid the homology modeling process and decrease the level of uncertainty by evaluating homology models to enrich known actives from decoys in docking experiments and/or to allow docking poses that match data from mutational studies (often termed 'ligand-based', 'ligand-guided', 'ligand-steered' or 'ligand-supported homology modeling'). Especially G-protein coupled receptors (GPCRs) were extensively studied using such approaches including serotonin receptors [4], dopamine receptors [5], GABA_B receptor [6] and neurokinin receptor 1 [7].

Most of these approaches heavily depend on scoring algorithms employed by docking programs to rank ligand poses and to estimate binding affinity [4–6]. However, docking scores often poorly corelate with binding affinity [8]. Also, searching for or optimizing a single homology model to bind a diverse set of ligands is arguable, since very different ligands might bind to or induce different protein conformations [9]. In contrast, Evers and Klebe avoided the use of docking scores by optimizing a homology model of the neurokinin receptor 1 to allow interactions with a single ligand that was extensively investigated including structure activity relationship of the ligand and mutational studies of the receptor to identify interacting amino acid chains [7]. Though, relying on mutational data can also be misleading, since mutations distant from the protein binding pocket can also drastically affect ligand binding [10].

In this study, we were interested if a single, yet important and reliable interaction can be exploited in a ligand-guided homology modeling workflow for the histamine H_3 receptor (H_3R) to gain structural knowledge about the binding site and to guide the selection of a homology model for subsequent virtual screening. We focused on an interaction of charged functional groups between ligands and aminergic GPCRs, which is well characterized and has been observed in multiple crystal structures of different GPCRs [11,12]. H_3R was selected as target for several reasons: (i) ligand data is publicly available, (ii) crystal structure is currently still missing, (iii) H_3R is an important drug target discussed for many severe diseases including Alzheimer's disease, schizophrenia, Parkinson's disease, narcolepsy, pain, and obesity among others [13,14] and (iv) a recent study of us revealed that H_3R and melanin-concentrating hormone receptor 1 can be inhibited by the same ligand which could be potentially used in obesity treatment [15]. In this project, 1000 homology models were generated and evaluated for allowing a charged interaction with a defined set of ligands. Best and worst performing models were structurally investigated and revealed the importance of distinct binding site residue conformations for proper ligand docking. The highest ranked model was used for a pharmacophore-based virtual screening campaign and led to the identification of two novel H_3R ligands with nanomolar affinity.

Results and discussion

Ligand-guided homology modeling

A template search revealed that the crystal structure of H_1R (3RZE [16]) does not show the highest sequence similarity to H_3R. Also, the extracellular loop 2 close to the orthosteric binding pocket is not resolved in the H_1R structure. Hence, homology modeling was performed with a multiple-template approach employing crystal structures of H_1R, muscarinic M_2 receptor (M_2R) and muscarinic M_3 (M_3R) receptor to generate 1000 homology models of H_3R with MODELLER 9.15 [17]. The average heavy atom RMSD of 1.2 Å was calculated with VMD 1.9.2 [18], whereat side chain heavy atoms were more flexible (1.6 Å) than backbone heavy atoms (0.4 Å). A set of 9 antagonists [19] (Table C in S1 File) was chosen to guide the selection of a homology model for later pharmacophore studies. We were specifically interested into this ligand series, since we found highly similar molecules active against the melanin-concentrating hormone receptor 1 (MCHR1) and dual antagonism of H_3R and MCHR1 might present a potential treatment option for obesity [15]. Additionally, these ligands are rather big showing Y-shaped conformations and thus should allow the selection of a homology model with an open binding pocket able to harbor diverse ligands. Subsequently, models were scored for presence of a charged interaction between the docked ligands and $D_{3.32}$ (numbering from Ballesteros-Weinstein numbering scheme [20]), that is known to be essential for ligand binding to aminergic GPCRs (Fig 1A) [11]. Docking and scoring have been performed twice to control for variations introduced by the docking algorithm (Fig B in S1 File). The highest
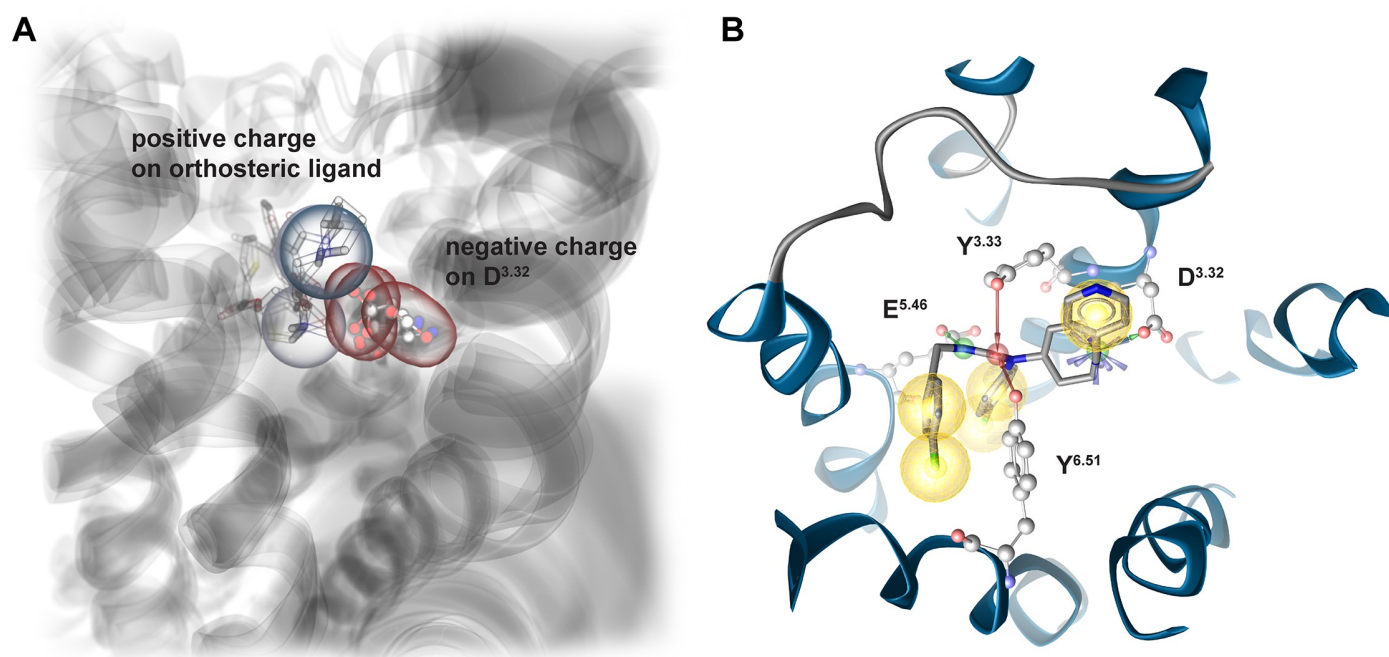
Fig 1. Ligand-guided homology modeling workflow exploits essential charged interaction known from aminergic GPCRs. (A) Aminergic GPCRs show a common charge interaction of highly diverse ligands with Aspartate 3.22 as illustrated for Eticlopride co-crystallized with the dopamine D3 receptor (3PBL [21]), Tiotropium co-crystallized with the muscarinic M_4 receptor (5DSG [22]) and Carazolol co-crystallized with the β_2 adrenoceptor (5JQH [23]). (B) Predominant binding mode of ligand series (Table C in S1 File) used for ligand-guided homology modeling. The depicted docking pose of CHEMBL1091834 involves a charged interaction with $D_{3.32}$, hydrogen bonds to $D_{3.32}$, $Y_{3.33}$, $E_{5.46}$ and $Y_{6.51}$ as well as several hydrophobic contacts. Red arrows–hydrogen bond acceptors, green arrows–hydrogen bond donors, blue star–positive ionizable, yellow sphere–hydrophobic contact.

https://doi.org/10.1371/journal.pone.0218820.g001

ranked model achieved an average score of 0.835 in both docking experiments. This means that 83.5% of the docking poses allow for a charged interaction with $D_{3.32}$. The predominant binding mode of docked ligands involves a charged interaction with $D_{3.32}$, hydrogen bonds with $D_{3.32}$, $Y_{3.33}$, $E_{5.46}$ and $Y_{6.51}$ as well as several hydrophobic contacts (Fig 1B). Interestingly, we found that 25% of generated models retrieved a score of 0.1 or lower. From these, 7 models had a score of 0, which means that none of the docking poses was involved in the essential charged interaction.

Thus, we got interested what determinants could be used to distinguish highly scored models from poorly scored models. First, 10 best and 10 worst performing models were tested for geometric errors like phi-psi outliers and heavy atom clashes in MOE 2015 [24] as well as with homology modeling evaluation programs including VERIFY 3D [25], ERRAT [26] and PROVE [27]. However, none of the applied methods led to a successful discrimination (Fig C in S1 File). Next, we analyzed structural differences by comparing the side chain atoms average position of 10 best and 10 worst performing models (Fig 2). The atom with the highest difference (4.7 Å) in the average position is a carboxyl oxygen of $E_{5.46}$ (Fig 2A). In the highly scored models $E_{5.46}$ is pointing inside the binding pocket (Fig 2B). This is in line with the predominant docking pose that is involved in a hydrogen bond with $E_{5.46}$. In contrast, poorly scored models show a conformation pointing outside the binding pocket. This conformation is also energetically unfavorable, since it is pointing toward the lipophilic membrane and no amino acid with opposite charge is present to compensate the negative charge. The importance of $E_{5.46}$ in ligand binding is in agreement with mutational studies [28] and was already described in previous homology modeling studies for H_3R [29,30]. Another atom with a rather high
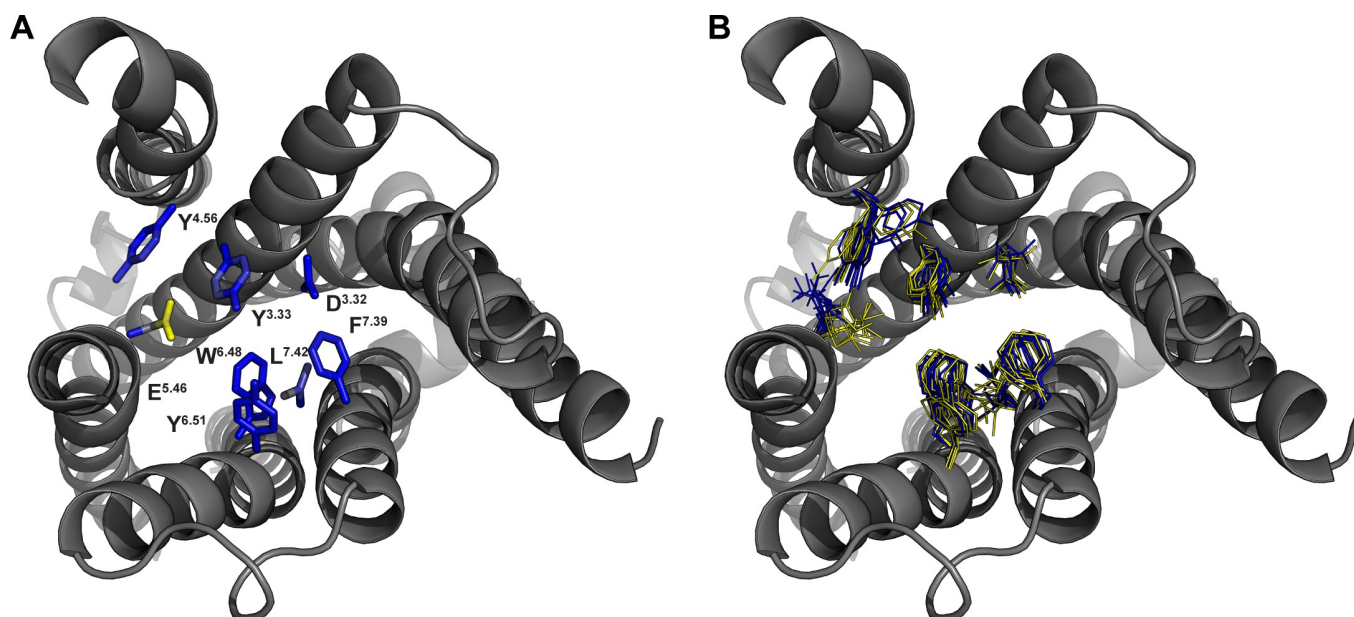
Fig 2. Best and worst scored homology models show distinct structural differences. Top view onto the orthosteric binding pocket of H₃R. Extracellular loop 2 is not shown for sake of clarity. (A) Structural differences were analyzed by calculating the difference in average side chain atom position of 10 highest and 10 lowest ranked models. Blue color indicates low difference, yellow high difference. (B) Sidechain conformations of 10 highest and 10 lowest ranked homology models. Yellow–high ranked models, blue–low ranked models.

https://doi.org/10.1371/journal.pone.0218820.g002

difference in mean atom position (2.1 Å) is a distal side chain carbon of $L_{7.42}$. However, we were not able to draw a clear connection to the docking results.

Virtual screening

The highest scored homology model was used for a screening campaign to identify novel H₃R ligands. 10 diverse antagonists (Table A in S1 File) were docked into the homology model. Constraints were added to focus on docking poses involved in interactions with the negatively charged carboxyl-group of $D^{3.32}$ and $E^{5.46}$, since all inverse agonists contain at least one positively charged group. Docking poses with favorable interaction patterns were found for only 5 out of 10 compounds and additionally analyzed to agree with published structure activity relationship. Derivatives of CHEMBL1923737 (Fig 3, model A) tolerate differently sized pyridone analogues indicating a location of the pyridone group outside the relatively narrow orthosteric binding pocket [31]. The literature about CHEMBL2151197 (Fig 3, model B) has only sparse structure active relationship data [32]. However, later pharmacophore modeling motivated us to include this docking pose in virtual screening. Analogues of CHEMBL2387294 show that 1 positively charged group can be exchanged by hydrophobic groups without loss of activity [33]. Hence, a docking pose was chosen that is extending outside the receptor with more space for different interactions (Fig 3, model C). Data for CHEMBL1269844 report a decrease in activity when attaching the naphthalene moiety in an extending fashion [34]. Concordantly, such molecule would lead to clashes with the receptor in the selected binding mode (Fig D part A in S1 File). The preferred docking pose of the histamine analogue CHEMBL214312 (Fig D part B in S1 File) is complexed between $D_{3.32}$ and $E_{5.46}$ [35]. This binding mode agrees well with several previous docking studies of histamine [29,30]. Each of the 5 chosen binding poses is involved in an interaction with charged residues $D_{3.32}$ and $E_{5.46}$, which is agreement with the common binding mode of aminergic GPCRs involving $D_{3.32}$ and with the importance of $E_{5.46}$
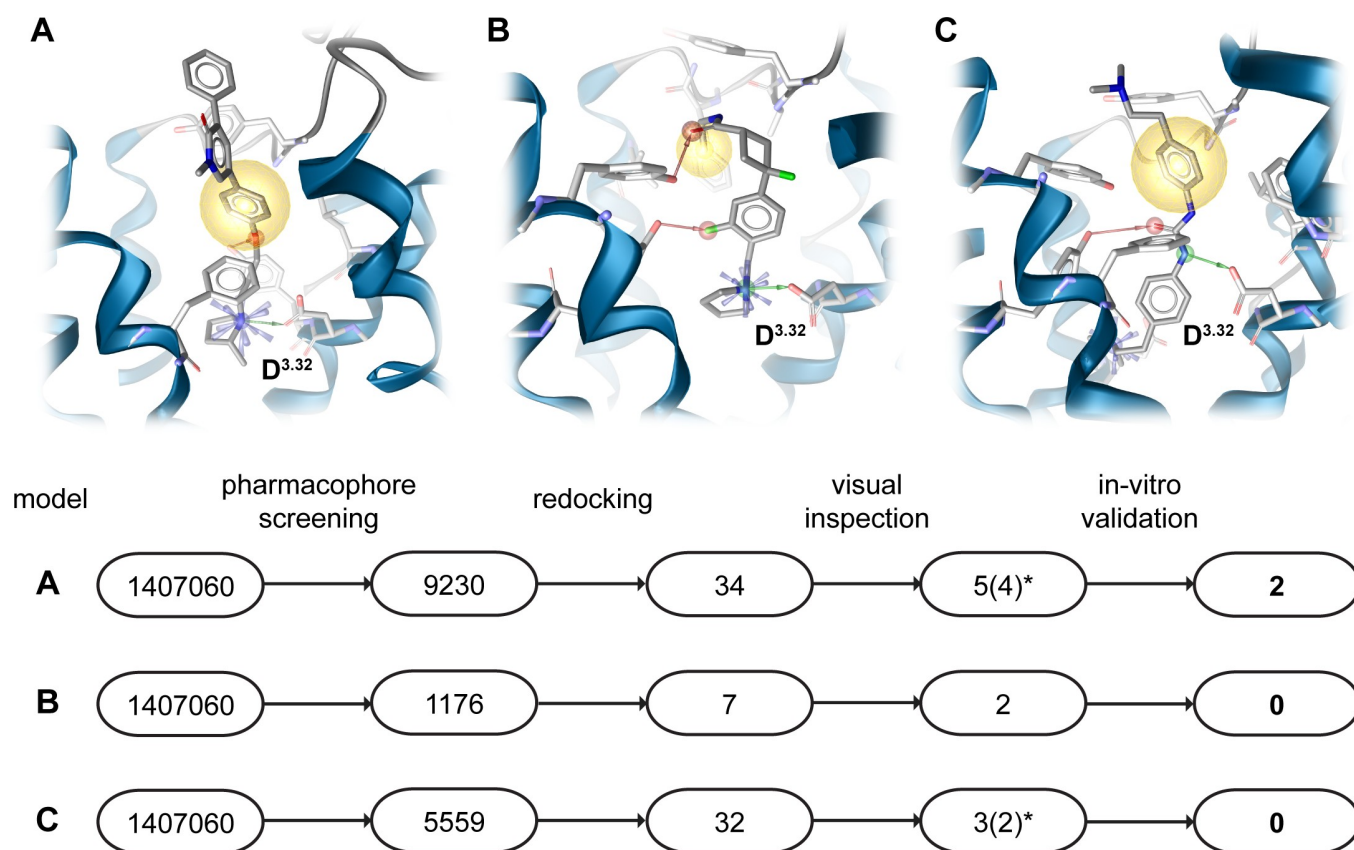
Fig 3. Virtual screening workflow results in 8 compounds out of 1.4 M for in-vitro validation. Workflow for virtual screening using 3 different pharmacophores based on docking poses of CHEMBL1923737 (model A), CHEMBL2151197 (model B) and CHEMBL2387294 (model C). Model A led to identification of compounds 3, 5, 6, 8, 9, model B to compounds 4, 7 and model C to compounds 1, 2, 10 (Fig 4, Table E in S1 File). * compounds 9 and 10 (model A and C) were removed from experimental testing due to insufficient purity as determined by LC-MS. Red arrows–hydrogen bond acceptors, green arrows–hydrogen bond donors, blue star–positive ionizable, yellow sphere–hydrophobic contact.

https://doi.org/10.1371/journal.pone.0218820.g003

for proper ligand placement in our homology modeling approach (Fig 2) that is further supported by mutational data [28] and previous docking studies [29,30]. Docking poses of CHEMBL1923737, CHEMBL2151197 (Fig 3, model A and B) and CHEMBL1269844 (Fig D part A in S1 File) only interact with $D_{3.32}$ despite the already described importance of $E_{5.46}$ in ligand binding. However, CHEMBL1923737 has only a single moiety able to act as hydrogen bond donor. Thus, it can only interact with one of such residues. Additionally, mutational data from the histamine H_1 receptor suggests that the amino acid at position 5.46 is only important for some ligands [36,37]. Selected complexes were minimized using SZYBKI [38] to allow binding site adaptation to the docked ligand. Pharmacophores were created and iteratively optimized using actives and property-matched decoys generated with DUD-E [39]. Three pharmacophores were found to efficiently discriminate between actives and decoys (Fig 3, Fig E in S1 File). Only pharmacophore model C includes interactions with residue $E_{5.46}$, whose conformation was found to be important for proper ligand docking in prior homology modeling selection. However, the 10 diverse inverse agonists used for this docking differ significantly from the shape of the Y-shaped compounds employed in ligand-guided homology modeling. Thus, it is not surprising that binding modes and interaction partners are to some extent different.

These pharmacophore models were used to screen a library of 1.4 M commercially available compounds (Enamine Ltd., Kyiv, Ukraine, www.enamine.net) resulting in almost 16,000 hits. The hits were docked into the respective minimized homology model and resulting docking poses were assessed for matching the previously screened pharmacophores. This procedure yielded 73 hits, which were visually inspected to identify hits complementing the receptor binding pocket surface. To broaden the chemical space of H_3R ligands, hits were also prioritized to cover positive ionizable head groups that are underrepresented or completely absent in the H_3R ligand data of the CHEMBL 20 database [40], i.e. terminal guanidino, 2,2,6,6-tetramethylpiperidino and secondary amino group (Fig 4). In total, 10 compounds were purchased for in-vitro testing. However, two compounds had to be excluded due to insufficient purity as determined by LC-MS (Table E in S1 File).

Two molecules (**5** and **6**) were found to bind H_3R in nanomolar concentration ranges (Fig 5). The identified binding mode indicates very similar interaction patterns including a charged interaction to $D_{3.32}$, hydrogen bonds to $D_{3.32}$ and $Y_{3.33}$ as well as several hydrophobic contacts. Moreover, we observed pi-cation interactions to $D_{3.32}$ and $Y_{3.33}$. Compound **6** shows an additional pi-cation interaction to $F_{7.39}$ which may contribute to its superior activity towards H_3R compared to compound **5**. Closest H_3R ligand analogues in CHEMBL 24 [40] were identified by employing Morgan fingerprints [41] implemented in RDKit [42] nodes for KNIME [43] with a Tanimoto score of 0.53 for compound **5** and of 0.36 for compound **6** (Fig 4). The closest analogues were characterized as inverse agonists indicating the same mode of action for the newly identified compounds **5** and **6** [44,45]. According to Morgan fingerprints [41] both compounds significantly differ from CHEMBL1923737 whose docking pose was used for pharmacophore modeling (Table D in S1 File). This is in line with frequently observed scaffold hopping in pharmacophore screening campaigns [46]. The thiazole motif of compound **5** has recently also been incorporated in new lead findings for this receptor subtype [47]. Compound **6** is known as CHEMBL1433079 and was tested in different high throughput bioassays. However, none of the reported primary screen activities was further investigated hindering a proper assessment of the data.

The remaining compounds bound H_3R at a concentration of 10 μM less than 50% and were not considered for in-depth activity characterization (Fig 4). Compounds **1–4** and **7** represent a molecule class that does not carry a lipophilic moiety (e.g. ethyl, cyclopropyl) at the charged head group like in compound **5** and **6** indicating an important role of this structural feature. Compound **8** does carry such hydrophobic moiety at the positively charged amine but was also found to be inactive. Hence, we speculate that the methyl group might be too small to effectively fulfill this structural role.

Conclusion

In this study, we successfully applied a ligand-guided homology modeling workflow to H_3R. Therefore, 1000 homology models were generated and evaluated for allowing a charged interaction in ligand docking experiments. A structural analysis of best and worst performing models revealed an important conformation of the binding site residue $E_{5.46}$ that is critical for proper ligand placement by the docking program. The best performing model was subsequently used in a virtual screening campaign and resulted in the identification of 2 novel H_3R ligands scaffolds with nanomolar affinity. Although successful, we do not claim that the best performing model is necessarily the most realistic one. However, we could show that many models were generated that allowed none or only few docking poses with the characteristic charged interaction. Thus, a single, easy-to-handle descriptor could be used to eliminate many low-quality homology models from further analysis.

Experimental section

Preparation of ligand data

The following workflow was conducted in KNIME [43] if not specified else. Histamine H_3 receptor (H_3R) ligand data was retrieved from Chembl 20 [40] database and filtered for
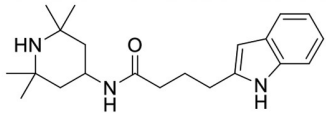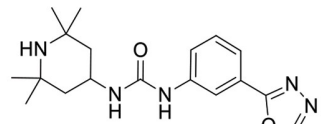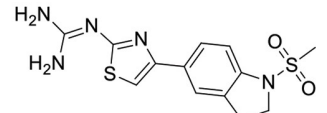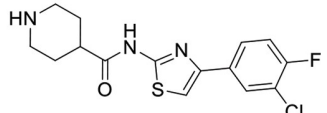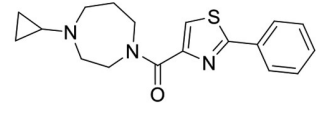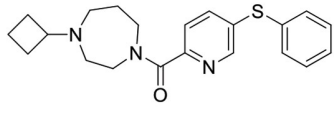
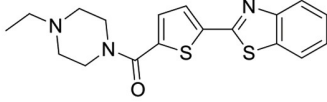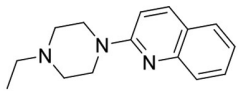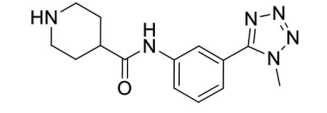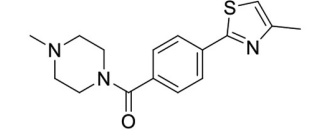| ID	structure	Binding at 10 μM [%]	K_i [nM]	Closest analogue in CHEMBL 24
1		17.6	-	-
2		1.2	-	-
3		-0.3	-	-
4		16.2	-	-
5		74.4	148.3	a
6		76.6	38.2	b
7		31.3	-	-
8		5.4	-	-

Fig 4. In-vitro validation of virtual screening hits identified 2 novel nanomolar H_3R ligands. Activity results of radioligand depletion assay against H_3R. K_i data is presented as mean values calculated from at least three independent experiments, each performed in triplicates. [a]CHEMBL1172076 with Tanimoto score of 0.53 when comparing with compound **5** using Morgan fingerprints, [b]CHEMBL180478 with Tanimoto score of 0.36 when comparing with compound **6** using Morgan fingerprints.

https://doi.org/10.1371/journal.pone.0218820.g004

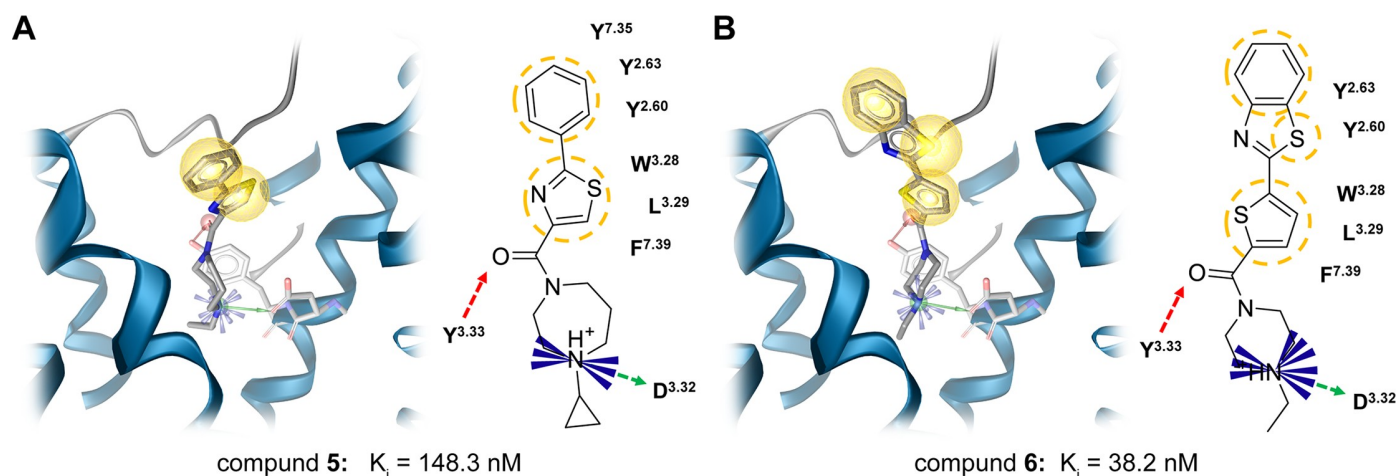**Fig 5. Potential binding modes of active ligands are very similar.** Observed interaction of screening hits **5** (A) and **6** (B). Red arrows—hydrogen bond acceptors, green arrows—hydrogen bond donors, blue star—positive ionizable, yellow sphere—hydrophobic contact.

molecular weight ($\leq$ 500 Da), confidence score (= 9), standard activity type (K_i, K_d, IC_{50} or EC_{50}), standard relation (=), standard activity value ($\leq$ 10) and standard activity unit (nM). Ligands with unclarified stereo centers were removed with a combination of RDKit [42] and Indigo [48] nodes. If multiple activities were available for a single ligand, binding data (K_i, K_d) was preferred over functional data IC_{50} or EC_{50}) and more recent data was preferred over older data. The literature of the remaining compounds was checked to remove agonists resulting in a final set of 632 inverse agonists. From this set 10 diverse inverse agonists (Table A in S1 File) were selected using the RDKit diversity picker based on MorganFeat fingerprints [41] (diameter = 4). This set was used for docking experiments to generate pharmacophores. Additionally, 100 diverse inverse agonists were selected for pharmacophore validation. Furthermore, the 100 diverse inverse agonists were used to generate decoys using the DUD-E decoy generator [39] for pharmacophore validation. The decoy set contains 3051 unique molecules. 3D coordinates of all molecules used in this study were generated and energetically minimized with the MMFF94s [49] force field using RDKit nodes[42]. Hydrogens were added, strong acids deprotonated and strong bases protonated by using the molecule wash function in MOE 2015 [24].

Homology modeling

The amino acid sequence of human H3R was retrieved from Uniprot [50] (Q9Y5N1) and employed for a homology model template search in the PDB[51] using the BLAST algorithm [52]. Structure files of the top ranked templates in the inactive conformation were used for an alignment in MOE 2015 [24]. Surprisingly, the crystal structure of H1R (3RZE [16]) did not show the highest sequence similarity to H3R. Also, the extracellular loop 2 (ECL2) close to the orthosteric binding pocket is not resolved in the H1R structure. Hence, homology modeling was performed with a multiple-template approach. MODELLER 9.15 [17] was used to generate 1000 homology models using H1R (3RZE [16]), muscarinic M2 receptor (M2R, 3UON [53]) and muscarinic M3 receptor (M3R, 4U15 [54]) as templates. Since ECL2 is not completely resolved in the H1R structure (3RZE), unresolved ECL2 parts were built by MODELLER solely based on M2R (3UON) and M3R (4U15). The sequence alignment as well as changed parameters of MODELLER functions can be found in the supporting information (Fig A and Table B in S1 File).

Docking experiments

A set of 9 inverse agonists [19] (Table C in S1 File) was chosen to guide the homology model selection and docked into all homology models using GOLD 5.2 [55] with default settings if not specified otherwise. The active site was defined by residues that are known from other aminergic GPCRs to be involved in ligand binding ($D_{3.32}$, $Y_{3.33}$, $Y_{4.56}$, $E_{5.46}$, $W_{6.48}$, $Y_{6.51}$ and $P_{7.39}$) [11]. 10 conformations were generated per molecule with the genetic algorithm set to 'Library Screening'. Early termination was disabled resulting in 90 conformations per homology model. Docking results were analyzed for ionic interaction between the ligand and $D_{3.32}$ that is characteristic for aminergic GPCRs [11]. Less or equal than 6 Å between the carbon atom of the carboxyl group of $D_{3.32}$ and the positively charged amine of the ligand was considered to be sufficient for ionic interaction. Docking and scoring have been performed to twice to control for variation introduced by the docking algorithm (Fig B in S1 File).

Homology model evaluation

10 best and 10 worst performing models were tested for geometric errors like phi-psi outliers and heavy atom clashes in MOE 2015 [24] as well as with homology modeling evaluation programs including VERIFY 3D [25], ERRAT [26] and PROVE [27]. No statistically significant difference was found (Fig C in S1 File).

Pharmacophore generation

The 10 diverse H_3R inverse agonists generated as described above were docked into the selected homology model using GOLD 5.2 [55] with default settings if not specified otherwise. The active site was defined by residues that are known from other aminergic GPCRs to be involved in ligand binding($D_{3.32}$, $Y_{3.33}$, $Y_{4.56}$, $E_{5.46}$, $W_{6.48}$, $Y_{6.51}$ and $P_{7.39}$) [11]. 10 conformations were generated per molecules with flip ring corners, flip pyramidal N and generate diverse solutions settings enabled and early termination setting disabled. Protein HBond constraints with a constraint weight of 10 and a minimum H-bond geometry weight of 0.005 were added to focus on conformations involving hydrogen bonds to carboxyl oxygens of $D^{3.32}$ and $E^{5.46}$, since all docked ligands contain a positively charged group that should interact with negatively charged carboxyl-group of $D^{3.32}$ or $E^{5.42}$. Docking results were analyzed in LigandScout 3.12 [56] for interactions explaining the structure-activity relationship. Selected complexes were minimized using Szybki 1.8.0.1 [38] with the MMFF94s forcefield and the Poisson-Boltzmann model. Sidechains within 10 Å were set flexible to allow adaption of the binding site residues to the docked ligand. LigandScout 3.12 was used to generate pharmacophores of the minimized complexes. Default pharmacophores generated with LigandScout 3.12 were optimized against a set of 100 diverse active inverse agonists and 3051 decoys by removing features or increasing the tolerance radius of selected features if supported by the structure activity relationship. Three pharmacophores were found to successfully discriminate between actives and decoys according to receiver operating characteristic curves (Fig E in S1 File).

Virtual screening and selection

The three selected pharmacophores were employed to screen a library of 1464080 molecules (Enamine Ltd., Kyiv, Ukraine, www.enamine.net) using LigandScout 3.12 [56] resulting in 15965 hits. These hits were redocked into the respective minimized model using GOLD 5.2 with default settings if not specified otherwise. The active site was defined by residues that are known from other aminergic GPCRs to be involved in ligand binding ($D_{3.32}$, $Y_{3.33}$, $Y_{4.56}$, $E_{5.46}$, $W_{6.48}$, $Y_{6.51}$ and $P_{7.39}$) [11]. 10 conformations were generated per molecules with flip ring

corners, flip pyramidal N and generate diverse solutions settings enabled. The redocked poses were scored to match the features of the respective pharmacophore model resulting in 73 hits. This set was visually inspected, and 10 molecules were selected for purchase. Ordered compounds were analyzed for purity with LC-MS leading to exclusion of 2 molecules from further analysis. The 8 remaining molecules possess purities of at least 95% and were tested in-vitro for activity against H_3R (Table E in S1 File).

In-vitro experiments

Radioligand depletion assays were performed as described previously using crude hH_3R membrane extracts obtained from HEK-293 cells stably expressing the hH_3R [15,57]. Briefly, crude membrane extracts were incubated with various concentrations of test ligands (between 0.01 nM and 100 μM) and [3H]-N-alpha-methylhistamine. Bound radioligand were harvested through GF/B filters and measured using liquid scintillation counting. Data analysis were performed with GraphPad Prism 6 using non-linear regression. The Ki values for each experiment were obtained according to Cheng-Prusoff and converted to pKi values to allow statistical analysis. Mean values were calculated from at least three independent experiments, each performed in triplicates (Table E in S1 File).

Supporting information

S1 File. PDF File with used molecular structures as well as more detailed parameters and results.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: David Schaller, Gerhard Wolber.

Investigation: David Schaller, Stefanie Hagenow, Holger Stark.

Methodology: David Schaller.

Resources: Gerhard Wolber.

Software: David Schaller, Gerhard Wolber.

Writing – original draft: David Schaller, Holger Stark.

Writing – review & editing: Holger Stark, Gerhard Wolber.

References

1. Lavecchia A, Giovanni C. Virtual Screening Strategies in Drug Discovery: A Critical Review. Curr Med Chem. 2013; 20: 2839–2860. https://doi.org/10.2174/09298673113209990001 PMID: 23651302

2. Ripphausen P, Nisius B, Peltason L, Bajorath J. Quo Vadis, Virtual Screening? A Comprehensive Survey of Prospective Applications. J Med Chem. 2010; 53: 8461–8467. https://doi.org/10.1021/jm101020z PMID: 20929257

3. Schmidt T, Bergner A, Schwede T. Modelling three-dimensional protein structures for applications in drug design. Drug Discov Today. 2014; 19: 890–897. https://doi.org/10.1016/j.drudis.2013.10.027 PMID: 24216321

4. Rodríguez D, Ranganathan A, Carlsson J. Strategies for improved modeling of GPCR-drug complexes: Blind predictions of serotonin receptors bound to ergotamine. J Chem Inf Model. 2014; 54: 2004–2021. https://doi.org/10.1021/ci5002235 PMID: 25030302

5. Kołaczkowski M, Bucki A, Feder M, Pawłowski M. Ligand-Optimized Homology Models of D 1 and D 2 Dopamine Receptors: Application for Virtual Screening. J Chem Inf Model. 2013; 53: 638–648. https://doi.org/10.1021/ci300413h PMID: 23398329

6. Freyd T, Warszycki D, Mordalski S, Bojarski AJ, Sylte I, Gabrielsen M. Ligand-guided homology modelling of the GABAB2 subunit of the GABAB receptor. PLoS One. 2017; 12: e0173889. https://doi.org/10.1371/journal.pone.0173889 PMID: 28323850

7. Evers A, Klebe G. Successful Virtual Screening for a Submicromolar Antagonist of the Neurokinin-1 Receptor Based on a Ligand-Supported Homology Model. J Med Chem. 2004; 47: 5381–5392. https://doi.org/10.1021/jm0311487 PMID: 15481976

8. Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, et al. A critical assessment of docking programs and scoring functions. J Med Chem. 2006; 49: 5912–31. https://doi.org/10.1021/jm050362n PMID: 17004707

9. Teague SJ. Implications of protein flexibility for drug discovery. Nat Rev Drug Discov. 2003; 2: 527–541. https://doi.org/10.1038/nrd1129 PMID: 12838268

10. Ragland DA, Whitfield TW, Lee S-K, Swanstrom R, Zeldovich KB, Kurt-Yilmaz N, et al. Elucidating the Interdependence of Drug Resistance from Combinations of Mutations. J Chem Theory Comput. 2017; 13: 5671–5682. https://doi.org/10.1021/acs.jctc.7b00601 PMID: 28915040

11. Michino M, Beuming T, Donthamsetti P, Newman AH, Javitch JA, Shi L. What can crystal structures of aminergic receptors tell us about designing subtype-selective ligands? Pharmacol Rev. 2015; 67: 198–213. https://doi.org/10.1124/pr.114.009944 PMID: 25527701

12. Nikolic K, Agbaba D, Stark H. Pharmacophore modeling, drug design and virtual screening on multi-targeting procognitive agents approaching histaminergic pathways. J Taiwan Inst Chem Eng. 2015; 46: 15–29. https://doi.org/10.1016/j.jtice.2014.09.017

13. Berlin M, Boyce CW, De Lera Ruiz M. Histamine H3 receptor as a drug discovery target. J Med Chem. 2011; 54: 26–53. https://doi.org/10.1021/jm100064d PMID: 21062081

14. Khanfar MA, Affini A, Lutsenko K, Nikolic K, Butini S, Stark H. Multiple Targeting Approaches on Histamine H3 Receptor Antagonists. Front Neurosci. 2016; 10: 1–17. https://doi.org/10.3389/fnins.2016.00001d>

15. Schaller D, Hagenow S, Alpert G, Naß A, Schulz R, Bermudez M, et al. Systematic Data Mining Reveals Synergistic H3R/MCHR1 Ligands. ACS Med Chem Lett. 2017; 8: 648–53. https://doi.org/10.1021/acsmedchemlett.7b00118 PMID: 28626527

16. Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, et al. Structure of the human histamine H1 receptor complex with doxepin. Nature. 2011; 475: 65–70. https://doi.org/10.1038/nature10236 PMID: 21697825

17. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinforma. 2016; 54: 5.6.1–5.6.37. https://doi.org/10.1002/cpbi.3 PMID: 27322406

18. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996; 14: 33–8, 27–8. https://doi.org/10.1016/0263-7855(96)00018-5 PMID: 8744570

19. Berlin M, Lee YJ, Boyce CW, Wang Y, Aslanian R, McCormick KD, et al. Reduction of hERG inhibitory activity in the 4-piperidinyl urea series of H3 antagonists. Bioorg Med Chem Lett. 2010; 20: 2359–64. https://doi.org/10.1016/j.bmcl.2010.01.121 PMID: 20188550

20. Ballesteros JA, Weinstein H. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. Methods Neurosci. 1995; 25: 366–428. https://doi.org/10.1016/S1043-9471(05)80049-7

21. Chien EYT, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, et al. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. Science. 2010; 330: 1091–5. https://doi.org/10.1126/science.1197410 PMID: 21097933

22. Thal DM, Sun B, Feng D, Nawaratne V, Leach K, Felder CC, et al. Crystal structures of the M1 and M4 muscarinic acetylcholine receptors. Nature. 2016; 531: 335–340. https://doi.org/10.1038/nature17188 PMID: 26958838

23. Staus DP, Strachan RT, Manglik A, Pani B, Kahsai AW, Kim TH, et al. Allosteric nanobodies reveal the dynamic range and diverse mechanisms of G-protein-coupled receptor activation. Nature. 2016; 535: 448–52. https://doi.org/10.1038/nature18636 PMID: 27409812

24. Chemical Computing Group Inc. Molecular Operating Environment (MOE). Montreal, QC, Canada; 2015. pp. 1010 Sherbooke St. West, Suite #910.

25. Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature. 1992; 356: 83–5. https://doi.org/10.1038/356083a0 PMID: 1538787

26. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993; 2: 1511–9. https://doi.org/10.1002/pro.5560020916 PMID: 8401235

27. Pontius J, Richelle J, Wodak SJ. Deviations from standard atomic volumes as a quality measure for protein crystal structures. J Mol Biol. 1996; 264: 121–36. https://doi.org/10.1006/jmbi.1996.0628 PMID: 8950272

28. Uveges AJ, Kowal D, Zhang Y, Spangler TB, Dunlop J, Semus S, et al. The role of transmembrane helix 5 in agonist binding to the human H3 receptor. J Pharmacol Exp Ther. 2002; 301: 451–8. Available: http://www.ncbi.nlm.nih.gov/pubmed/11961043 https://doi.org/10.1124/jpet.301.2.451 PMID: 11961043

29. Kiss R, Keserű GM. Structure-based discovery and binding site analysis of histamine receptor ligands. Expert Opin Drug Discov. Taylor & Francis; 2016; 11: 1165–1185. https://doi.org/10.1080/17460441.2016.1245288 PMID: 27704986

30. Kooistra AJ, Kuhne S, de Esch IJP, Leurs R, de Graaf C. A structural chemogenomics analysis of aminergic GPCRs: lessons for histamine receptor ligand design. Br J Pharmacol. 2013; 170: 101–126. https://doi.org/10.1111/bph.12248 PMID: 23713847

31. Becknell NC, Lyons JA, Aimone LD, Gruner JA, Mathiasen JR, Raddatz R, et al. Synthesis and evaluation of pyridone-phenoxypropyl-R-2-methylpyrrolidine analogues as histamine H3 receptor antagonists. Bioorg Med Chem Lett. 2011; 21: 7076–80. https://doi.org/10.1016/j.bmcl.2011.09.091 PMID: 22014551

32. Wager TT, Pettersen BA, Schmidt AW, Spracklin DK, Mente S, Butler TW, et al. Discovery of two clinical histamine H(3) receptor antagonists: trans-N-ethyl-3-fluoro-3-[3-fluoro-4-(pyrrolidinylmethyl)phenyl] cyclobutanecarboxamide (PF-03654746) and trans-3-fluoro-3-[3-fluoro-4-(pyrrolidin-1-ylmethyl)phenyl]-N-(2-methylpropyl)cyclobuta. J Med Chem. 2011; 54: 7602–20. https://doi.org/10.1021/jm200939b PMID: 21928839

33. Gao Z, Hurst WJ, Guillot E, Czechtizky W, Lukasczyk U, Nagorny R, et al. Discovery of aryl ureas and aryl amides as potent and selective histamine H3 receptor antagonists for the treatment of obesity (Part I). Bioorganic Med Chem Lett. Elsevier Ltd; 2013; 23: 3416–3420. https://doi.org/10.1016/j.bmcl.2013.03.080 PMID: 23591110

34. Anderson JT, Campbell M, Wang J, Brunden KR, Harrington JJ, Stricker-Krongrad A, et al. Investigation of 4-piperidinols as novel H3 antagonists. Bioorg Med Chem Lett. 2010; 20: 6246–9. https://doi.org/10.1016/j.bmcl.2010.08.099 PMID: 20833043

35. Watanabe M, Kazuta Y, Hayashi H, Yamada S, Matsuda A, Shuto S. Stereochemical diversity-oriented conformational restriction strategy. Development of potent histamine H3 and/or H4 receptor antagonists with an imidazolylcyclopropane structure. J Med Chem. 2006; 49: 5587–96. https://doi.org/10.1021/jm0603318 PMID: 16942032

36. Moguilevsky N, Varsalona F, Guillaume JP, Noyer M, Gillard M, Daliers J, et al. Pharmacological and functional characterisation of the wild-type and site-directed mutants of the human H1 histamine receptor stably expressed in CHO cells. J Recept Signal Transduct Res. 15: 91–102. https://doi.org/10.3109/10799899509045210 PMID: 8903934

37. Bruysters M, Pertz HH, Teunissen A, Bakker RA, Gillard M, Chatelain P, et al. Mutational analysis of the histamine H1-receptor binding pocket of histaprodifens. Eur J Pharmacol. 2004; 487: 55–63. https://doi.org/10.1016/j.ejphar.2004.01.028 PMID: 15033376

38. SZYBKI 1.8.0.1: OpenEye Scientific Software, Santa Fe, NM; http://www.eyesopen.com.

39. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. J Med Chem. 2012; 55: 6582–6594. https://doi.org/10.1021/jm300687e PMID: 22716043

40. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: An update. Nucleic Acids Res. 2014; 42. https://doi.org/10.1093/nar/gkt1031 PMID: 24214965

41. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010; 50: 742–754. https://doi.org/10.1021/ci100050t PMID: 20426451

42. RDKit: Open-source cheminformatics; http://www.rdkit.org.

43. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. KNIME: The Konstanz Information Miner. 2008. pp. 319–326. https://doi.org/10.1007/978-3-540-78246-9_38

44. Zaragoza F, Stephensen H, Peschke B, Rimvall K. 2-(4-Alkylpiperazin-1-yl)quinolines as a New Class of Imidazole-Free Histamine H 3 Receptor Antagonists. J Med Chem. 2005; 48: 306–311. https://doi.org/10.1021/jm040873u PMID: 15634025

45. Letavic MA, Aluisio L, Atack JR, Bonaventure P, Carruthers NI, Dugovic C, et al. Pre-clinical characterization of aryloxypyridine amides as histamine H3 receptor antagonists: Identification of candidates for clinical development. Bioorg Med Chem Lett. 2010; 20: 4210–4214. https://doi.org/10.1016/j.bmcl.2010.05.041 PMID: 20561786

46. Schneider Neidhart, Giller Schmid. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. Angew Chem Int Ed Engl. 1999; 38: 2894–2896. Available: http://www.ncbi.nlm.nih.gov/pubmed/10540384 PMID: 10540384

47. Khanfar MA, Reiner D, Hagenow S, Stark H. Design, synthesis, and biological evaluation of novel oxadiazole- and thiazole-based histamine H 3 R ligands. Bioorg Med Chem. Elsevier; 2018; 26: 4034–4046. https://doi.org/10.1016/j.bmc.2018.06.028 PMID: 29960729

48. Pavlov D, Rybalkin M, Karulin B, Kozhevnikov M, Savelyev A, Churinov A. Indigo: universal cheminformatics API. J Cheminform. 2011; 3: P4. https://doi.org/10.1186/1758-2946-3-S1-P4

49. Halgren TA. MMFF VI. MMFF94s option for energy minimization studies. J Comput Chem. 1999; 20: 720–729. https://doi.org/10.1002/(SICI)1096-987X(199905)20:7<720::AID-JCC7>3.0.CO;2-X

50. Consortium UniProt. UniProt: a hub for protein information. Nucleic Acids Res. 2015; 43: D204–12. https://doi.org/10.1093/nar/gku989 PMID: 25348405

51. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28: 235–42. https://doi.org/10.1093/nar/28.1.235 PMID: 10592235

52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

53. Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, et al. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. Nature. 2012; 482: 547–51. https://doi.org/10.1038/nature10753 PMID: 22278061

54. Thorsen TS, Matt R, Weis WI, Kobilka BK. Modified T4 Lysozyme Fusion Proteins Facilitate G Protein-Coupled Receptor Crystallogenesis. Structure. 2014; 22: 1657–64. https://doi.org/10.1016/j.str.2014.08.022 PMID: 25450769

55. Cole J, Willem M. Nissink J, Taylor R. Protein-Ligand Docking and Virtual Screening with GOLD. In: Alvarez J, Shoichet B, editors. Virtual Screening in Drug Discovery. Boca Raton: Taylor & Francis CRC Press; 2005. pp. 379–415. https://doi.org/10.1201/9781420028775.ch15

56. Wolber G, Langer T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model. 2005; 45: 160–9. https://doi.org/10.1021/ci049885e PMID: 15667141

57. Kottke T, Sander K, Weizel L, Schneider EH, Seifert R, Stark H. Receptor-specific functional efficacies of alkyl imidazoles as dual histamine H3/H4 receptor ligands. Eur J Pharmacol. 2011; 654: 200–8. https://doi.org/10.1016/j.ejphar.2010.12.033 PMID: 21237145

Supporting Information

Ligand-guided homology modeling drives identification of novel ligands of histamine H3 receptor

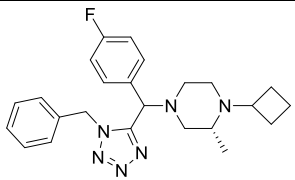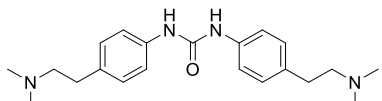David Schaller, Stefanie Hagenow, Holger Stark and Gerhard Wolber
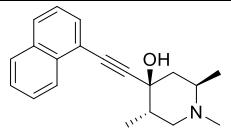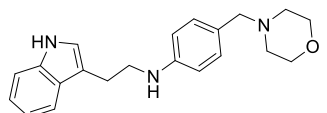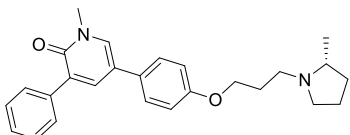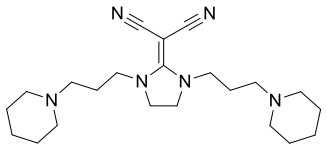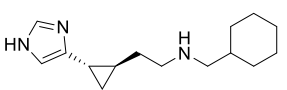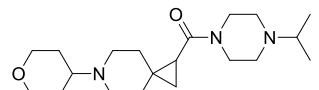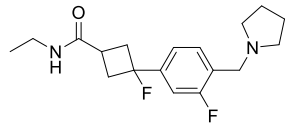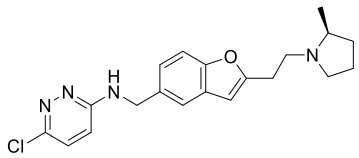
Table of contents

1. Computational experiments

1.1 Preparation of ligand data

Table A: Diverse H₃R ligands. 10 diverse ligands used for docking experiments in selected homology model.

CHEMBL1222946, IC$_{50}$=2.2 nM[1]

CHEMBL2387294, IC$_{50}$=6.5 nM[2]

CHEMBL1269844, IC$_{50}$=9 nM[3]

CHEMBL2419581, K$_i$=4.3 nM[4]

CHEMBL1923737, K$_i$=3.5 nM[5]

CHEMBL257391, K$_i$=2.4 nM[6]

CHEMBL214312, K$_i$=5.3 nM[7]

CHEMBL3124968, IC$_{50}$=8.3 nM[8]

CHEMBL2151197, K$_i$=2.3 nM[9]

CHEMBL362952, K$_i$=0.3 nM[10]

1.2 Homology modeling

```
                <-----------H1------------>      <-------------H2-----------
H3R        35   AVLAALMALLIVATVLGNALVMLAFVADSSLRTQNNFFLLNLAISDFLVGAFCIPLYVPY   94
H1R (3RZE)      MPLVVVLSTICLVTVGLNLLVLYAVRSERKLHTVGNLYIVSLSVADLIVGAVVMPMNILY
M2R (3UON)      VFIVLVAGSLSLVTIIGNILVMVSIKVNRHLQTVNNYFLFSLACADLIIGVFSMNLYTLY
M3R (4U15)      VFIAFLTGFLALVTIIGNILVIVAFKVNKQLKTVNNYFLLSLACADLIIGVISMNLFTTY
                 :.  :  .  : :.*: * **: :.   : *:*  * ::..*: :*::*.. : :    *
                                                     <ICL1>
```

```
                ->      <---------------H3---------------->          <-----
H3R        95   VLTGRWTFGRGLCKLWLVVDYLLCTSSAFNIVLISYDRFLSVTRAVSYRAQQGDTRRAVR  154
H1R (3RZE)      LLMSKWSLGRPLCLFWLSMDYVASTASIFSVFILCIDRYRSVQQPLRYL-KYRTKTRASA
M2R (3UON)      TVIGYWPLGPVVCDLWLALDYVVSNASVMNLLIISFDRYFCVTKPLTYP-VKRTTKMAGM
M3R (4U15)      IIMNRWALGNLACDLWLSIDYVASNASVMNLLVISFDRYFSITRPLTYR-AKRTTKRAGV
                 :   * :*    * :** :**: ..:* :.:.::. **: .: : : *    ...  *
                <ECL1>                                          <--ICL2--->
```

```
                ------H4---------------->             <------------H5---
H3R        155  KMLLVWVLAFLLYGPAILSWEYLSGGSSIPEGHCYAEFFYNWYFLITASTLEFFTPFLSV  214
H1R (3RZE)      TILGAWFLSF-LWVIPILGWNH/----/RREDKCETDFYDVTWFKVMTAIINFYLPTLLM
M2R (3UON)      MIAAAWVLSFILWAPAILFWQFIVGVRTVEDGECYIQFFSNAAVTFGTAIAAFYLPVIIM
M3R (4U15)      MIGLAWVISFVLWAPAILFWQYFVGKRTVPPGECFIQFLSEPTITFGTAIAAFYMPVTIM
                 :  .*.::* *:   ** *:.       .* :*   . . :: *: *   :
                                          <-----ECL2------>
```

```
                ---------->      <--------------H6---------------->    <----
H3R        215  TFFNLSIYLNIQR/LSRDRKVAKSLAVIVSIFGLCWAPYTLLMIIRAACHGHCVPDYWYE  395
H1R (3RZE)      LWFYAKIYKAVRQ/MNRERKAAKQLGFIMAAFILCWIPYFIFFMVIAFCKN-CCNEHLHM
M2R (3UON)      TVLYWHISRASKS/PSREKKVTRTILAILLAFIITWAPYNVMVLINTFCAP-CIPNTVWT
M3R (4U15)      TILYWRIYKETEK/LIKEKKAAQTLSAILLAFIITWTPYNIMVLVNTFCDS-CIPKTYWN
                 :   *     .    :::*.:: :   *:  *  : * ** ::.:: : *   * .
                                  <ICL3>                       <-ECL3->
```

```
                -------H7----------> <---H8--->
H3R        396  TSFWLLWANSAVNPVLYPLCHHSFRRAFTKL  426
H1R (3RZE)      FTIWLGYINSTLNPLIYPLCNENFKKTFKRI
M2R (3UON)      IGYWLCYINSTINPACYALCNATFKKTFKHL
M3R (4U15)      LGYWLCYINSTVNPVCYALCNKTFRTTFKTL
                 ** : **::** * **. .*: :*. :
```
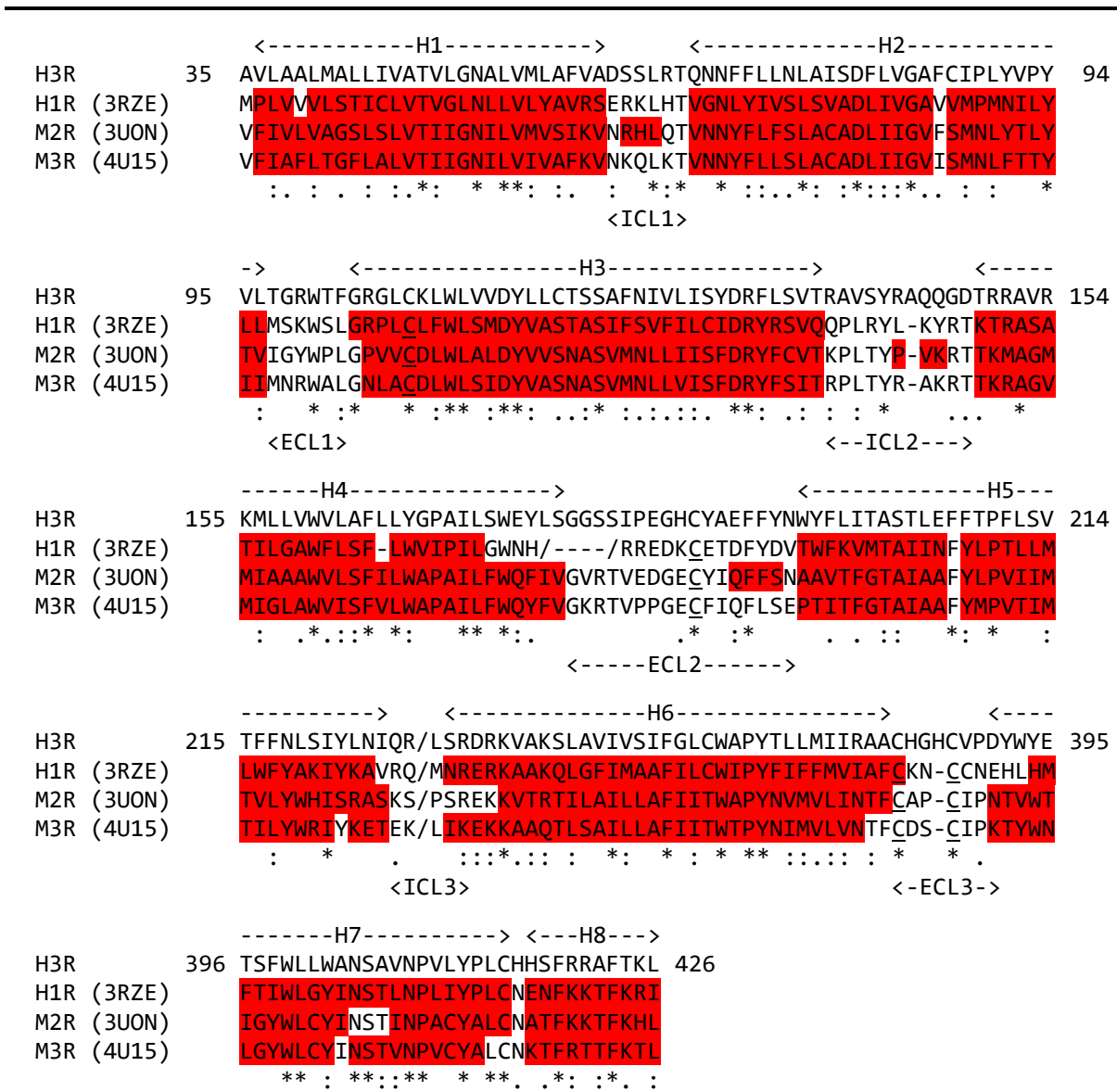
Fig A: Sequence alignment. The depicted multiple sequence alignment was used for ligand-guided homology modeling of human H₃R. The template sequences are illustrated according to structural properties. Red sections represent helices and underlined cysteines are involved in a disulfide bond. Furthermore, the multiple sequence alignment contains information about the naming of helices and loops and the sequence similarity. (H1) - helix 1, (ILC1) – intracellular loop 1, (ECL1) – extracellular loop 2, (*) - identical residues, (:) - residues with high similarity, (.) - residues with low similarity.

Table B: MODELLER parameters. Changed parameters of environ and allhmodel classes in homology modelling using MODELLER 9.15[11].

class	parameter	setting
environ	schedule_scale	physical.values(default=1.0, soft_sphere=0.7)
allhmodel	library_schedule	autosched.slow
allhmodel	max_var_iterations	300
allhmodel	md_level	refine.slow
allhmodel	repeat_optimization	2
allhmodel	max_molpdf	1e6

1.3 Docking experiments

Table C: Homology modeling ligand series. Structures of molecules used for guiding homology model selection. Provided activities are taken from the literature[12].
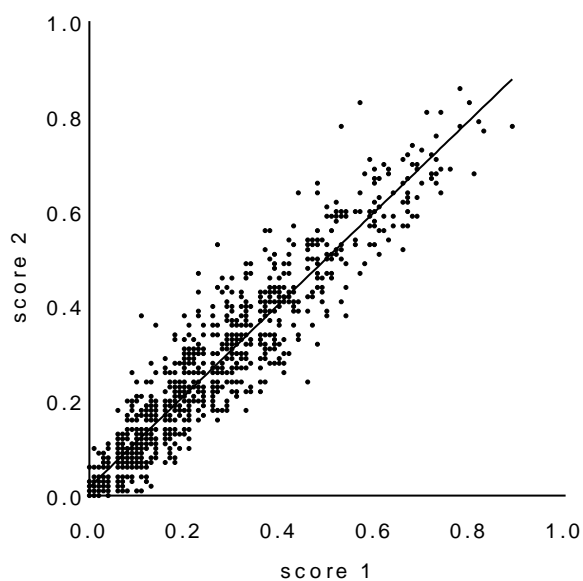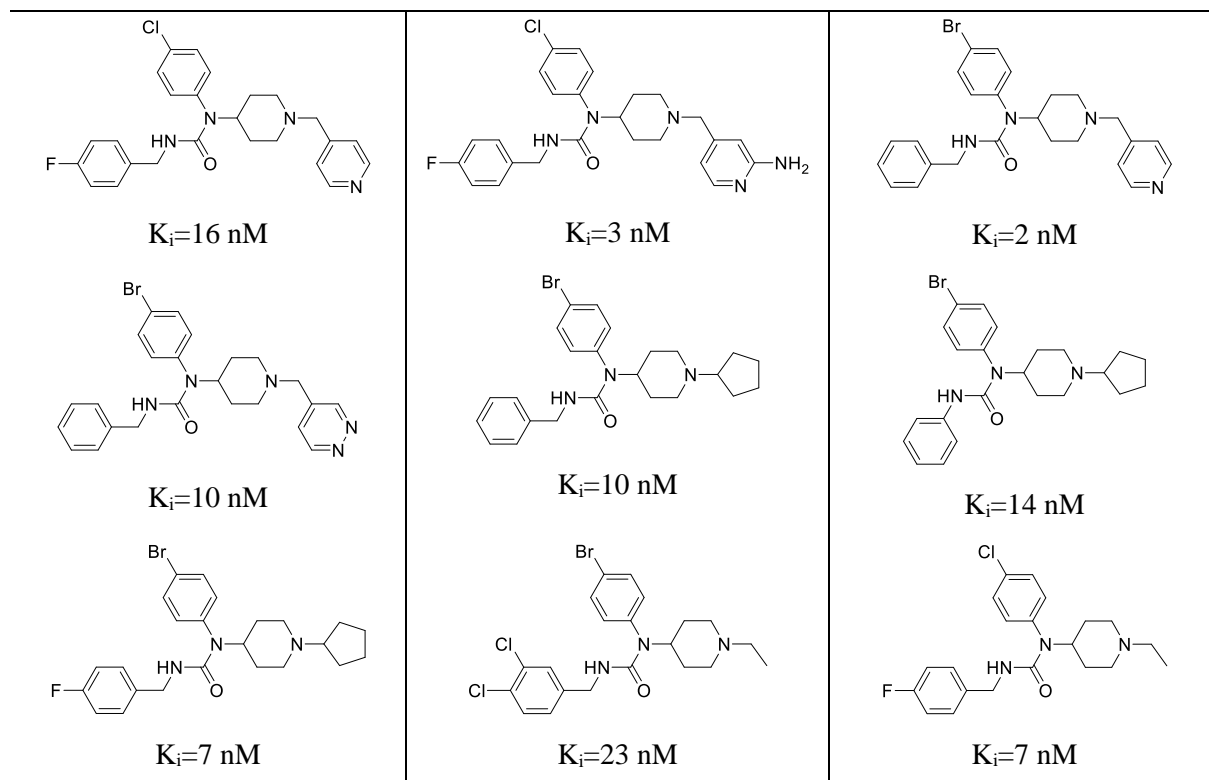

K_i=16 nM	K_i=3 nM	K_i=2 nM
K_i=10 nM	K_i=10 nM	K_i=14 nM
K_i=7 nM	K_i=23 nM	K_i=7 nM

Fig B: Scoring repeats. Scoring results from two independent docking experiments with linear regression line (R^2 = 0.89).
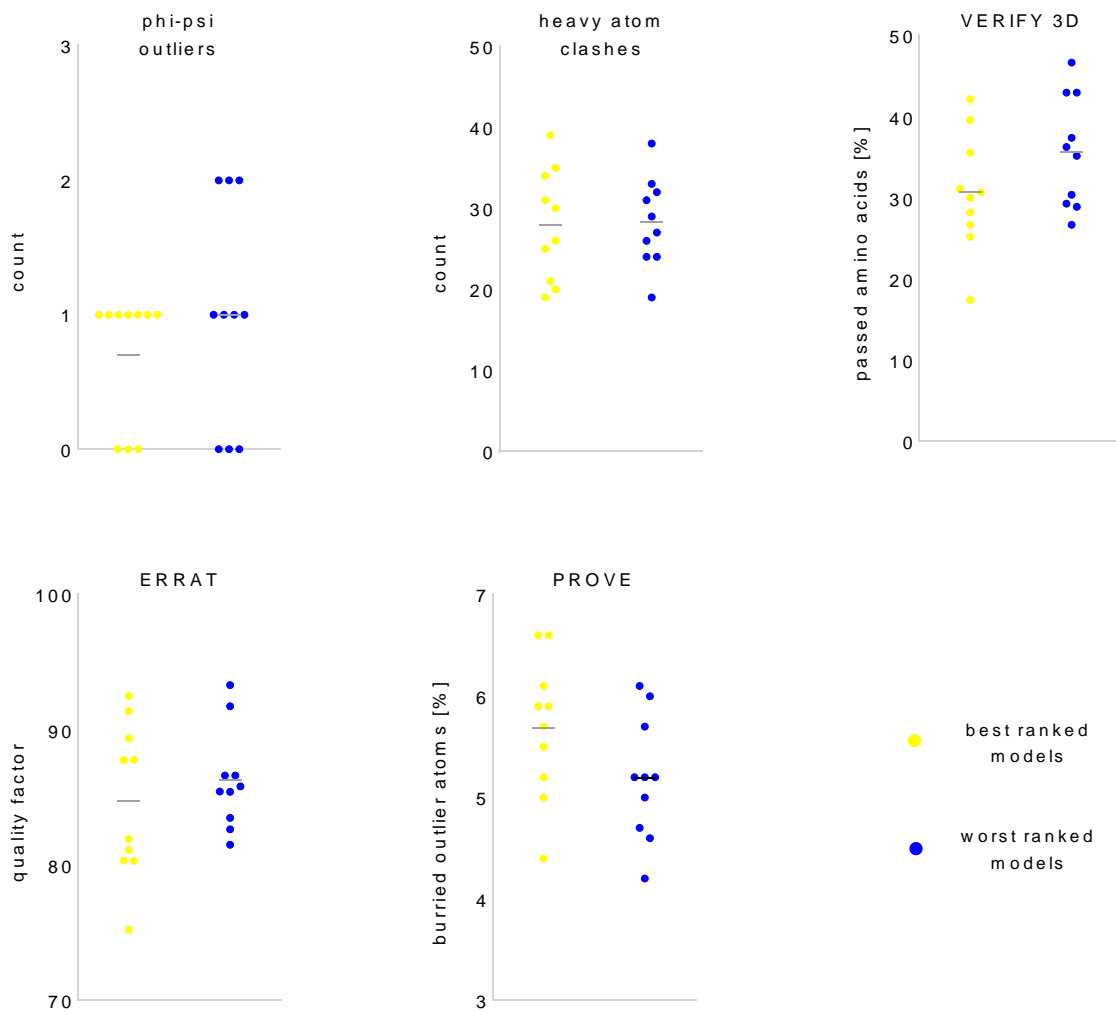
1.4 Homology model evaluation

Fig C: Homology model evaluation. Model quality comparison of 10 best and 10 worst performing models using Mann-Whitney U test as implemented in GraphPad Prism 6.

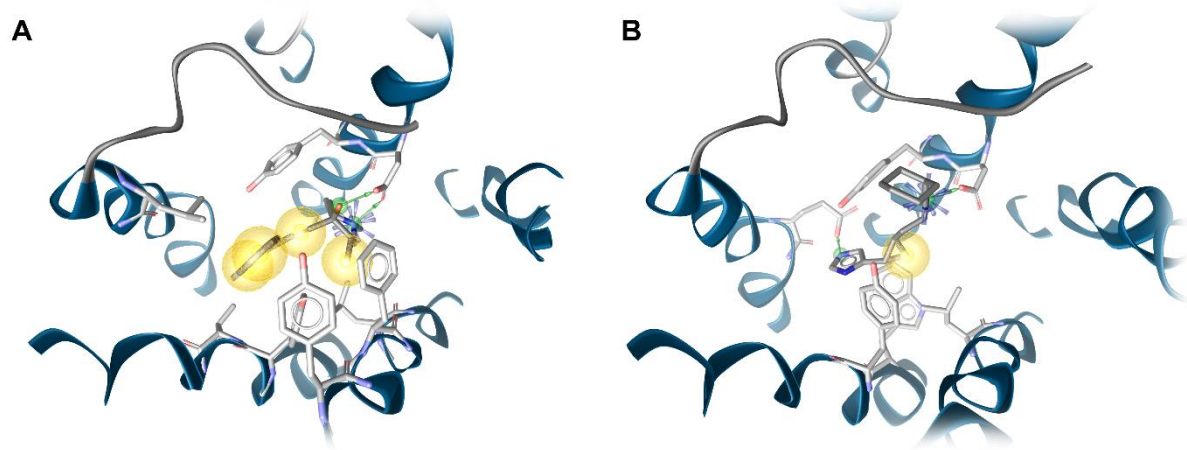## 1.5 Additional binding modes not used for pharmacophore screening

Fig D: Additional proposed binding modes of docked diverse antagonists for CHEMBL1269844[3] and CHEMBL214312[7]. Green arrows – hydrogen bond donors, blue star – positive ionizable, yellow sphere – hydrophobic contact.

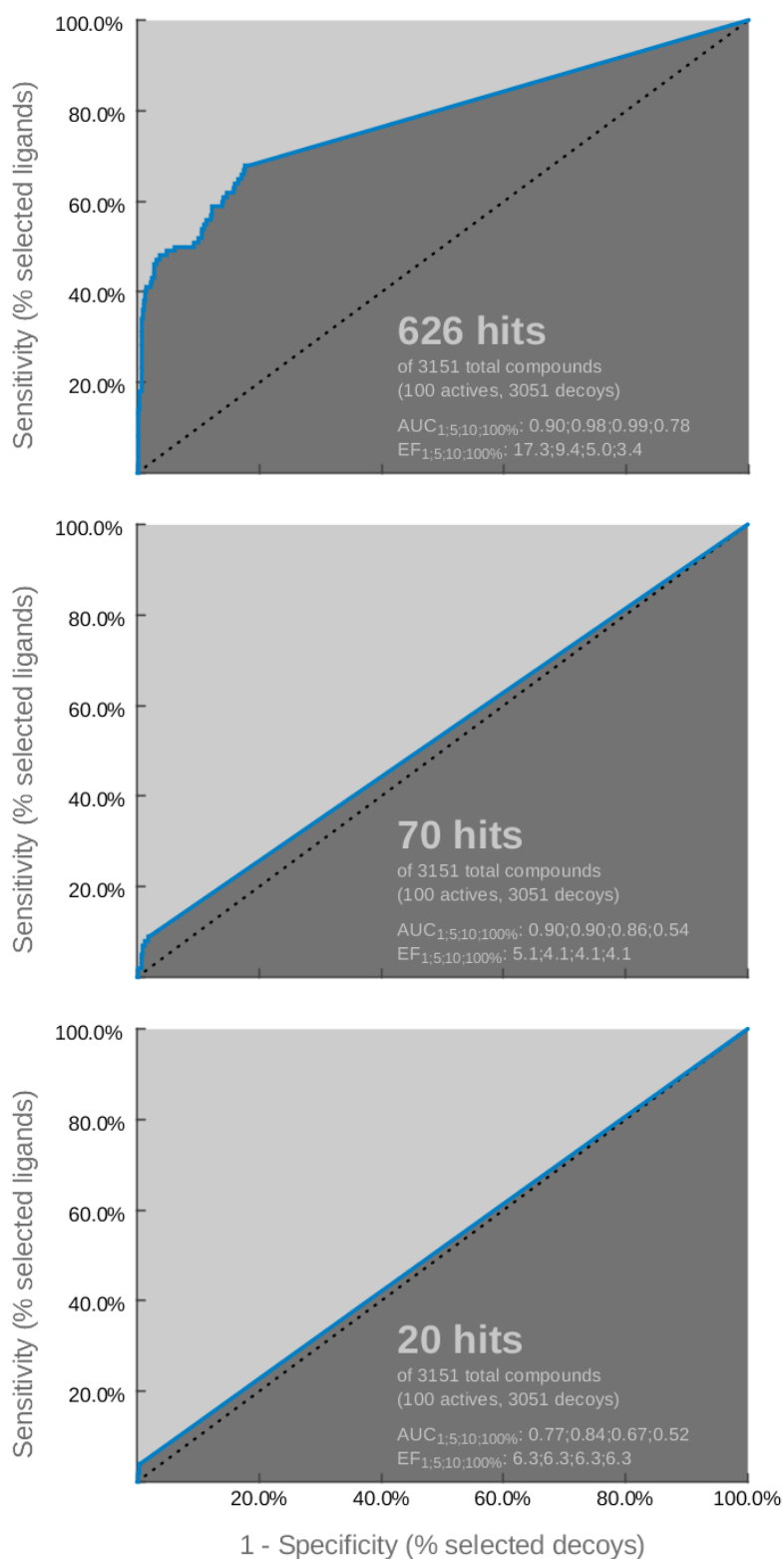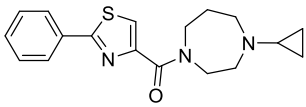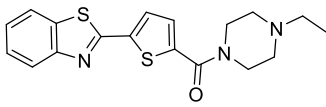## 1.6 ROC plot analysis of pharmacophores used for virtual screening
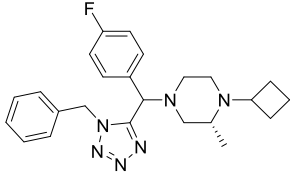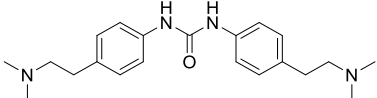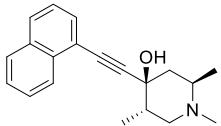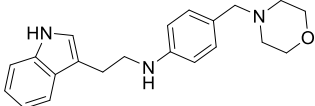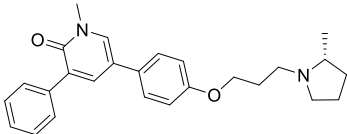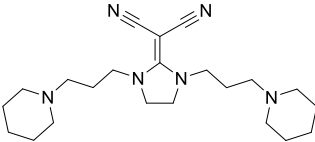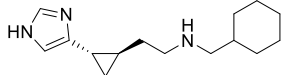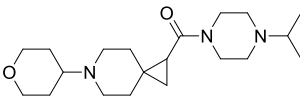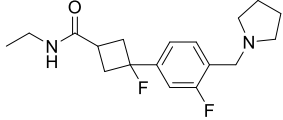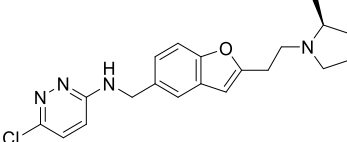
Fig E: ROC plot analysis of pharmacophores for virtual screening. (Top) Model A, CHEMBL1923737, (Middle) Model B, CHEMBL2151197, (Bottom) CHEMBL2387294.
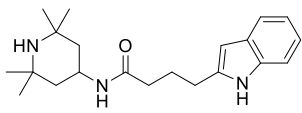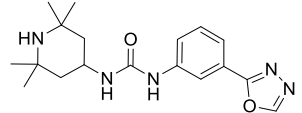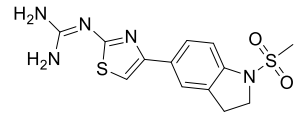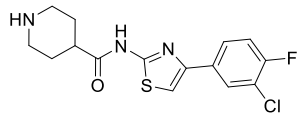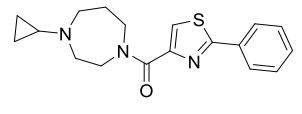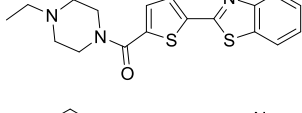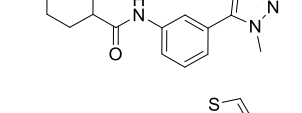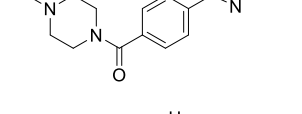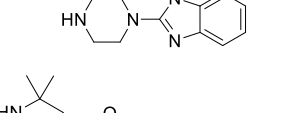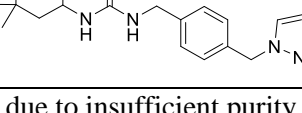
1.6 Similarity matrix

Table D: Similarity matrix between 10 diverse already known H₃R ligands and 2 newly identified H₃R ligands. Similarities are represented as Tanimoto scores calculated with Morgan fingerprints[13] as implanted in RDKit[14] nodes in KNIME[15].

	0.190	0.161
	0.109	0.125
	0.149	0.147
	0.127	0.139
	0.188	0.129
	0.09	0.121
	0.139	0.107
	0.200	0.118
	0.139	0.167
	0.151	0.136

2. In-vitro experiments

Table E: In-vitro results and purity. Purity and affinity of tested compounds analyzed with LC-MS and radioligand depletion experiment, respectively.

ID	structure	purity [%]	MW [g/mol]	m/z [M+H⁺]	Binding at 10 μM [%]	pKᵢ $\overline{x} \pm$ SEM
1		> 95	341.2	342.3	17.6	-
2		> 95	343.2	344.3	1.2	-
3		> 95	337.1	338.0	-0.3	-
4		> 95	339.1	340.0	16.2	-
5		> 95	327.1	328.0	74.4	**6.83 ± 0.23**
6		> 95	357.1	358.0	76.6	**7.46 ± 0.11**
7		> 95	286.2	287.3	31.3	-
8		> 95	301.1	302.0	5.4	-
9		**< 90**	202.1	203.0	[a]	
10		**< 90**	369.3	370.3	[a]	

[a] Excluded due to insufficient purity.

3. References

(1)   Davenport, A. J.; Stimson, C. C.; Corsi, M.; Vaidya, D.; Glenn, E.; Jones, T. D.; Bailey, S.; Gemkow, M. J.; Fritz, U.; Hallett, D. J. Discovery of Substituted Benzyl Tetrazoles as Histamine H3 Receptor Antagonists. *Bioorg. Med. Chem. Lett.* **2010**, *20* (17), 5165–5169.

(2)   Gao, Z.; Hurst, W. J.; Guillot, E.; Czechtizky, W.; Lukasczyk, U.; Nagorny, R.; Pruniaux, M. P.; Schwink, L.; Sanchez, J. A.; Stengelin, S.; Tang, L.; Winkler, I.; Hendrix, J. a.; George, P. G. Discovery of Aryl Ureas and Aryl Amides as Potent and Selective Histamine H3 Receptor Antagonists for the Treatment of Obesity (Part I). *Bioorganic Med. Chem. Lett.* **2013**, *23* (11), 3416–3420.

(3)   Anderson, J. T.; Campbell, M.; Wang, J.; Brunden, K. R.; Harrington, J. J.; Stricker-Krongrad, A.; Song, J.; Doucette, C.; Murphy, S.; Bennani, Y. L. Investigation of 4-Piperidinols as Novel H3 Antagonists. *Bioorg. Med. Chem. Lett.* **2010**, *20* (21), 6246–6249.

(4)   Tang, L.; Zhao, L.; Hong, L.; Yang, F.; Sheng, R.; Chen, J.; Shi, Y.; Zhou, N.; Hu, Y. Design and Synthesis of Novel 3-Substituted-Indole Derivatives as Selective H3 Receptor Antagonists and Potent Free Radical Scavengers. *Bioorg. Med. Chem.* **2013**, *21* (19), 5936–5944.

(5)   Becknell, N. C.; Lyons, J. A.; Aimone, L. D.; Gruner, J. A.; Mathiasen, J. R.; Raddatz, R.; Hudkins, R. L. Synthesis and Evaluation of Pyridone-Phenoxypropyl-R-2-Methylpyrrolidine Analogues as Histamine H3 Receptor Antagonists. *Bioorg. Med. Chem. Lett.* **2011**, *21* (23), 7076–7080.

(6)   Sasho, S.; Seishi, T.; Kawamura, M.; Hirose, R.; Toki, S.; Shimada, J. Diamine Derivatives Containing Imidazolidinylidene Propanedinitrile as a New Class of Histamine H3 Receptor Antagonists. Part I. *Bioorg. Med. Chem. Lett.* **2008**, *18* (7), 2288–2291.

(7)   Watanabe, M.; Kazuta, Y.; Hayashi, H.; Yamada, S.; Matsuda, A.; Shuto, S. Stereochemical Diversity-Oriented Conformational Restriction Strategy. Development of Potent Histamine H3 and/or H4 Receptor Antagonists with an Imidazolylcyclopropane Structure. *J. Med. Chem.* **2006**, *49* (18), 5587–5596.

(8)   Brown, D. G.; Bernstein, P. R.; Griffin, A.; Wesolowski, S.; Labrecque, D.; Tremblay, M. C.; Sylvester, M.; Mauger, R.; Edwards, P. D.; Throner, S. R.; Folmer, J. J.; Cacciola, J.; Scott, C.; Lazor, L. A.; Pourashraf, M.; Santhakumar, V.; Potts, W. M.; Sydserff, S.; Giguère, P.; Lévesque, C.; Dasser, M.; Groblewski, T. Discovery of Spirofused Piperazine and Diazepane Amides as Selective Histamine-3 Antagonists with in Vivo Efficacy in a Mouse Model of Cognition. *J. Med. Chem.* **2014**, *57* (3), 733–758.

(9)   Wager, T. T.; Pettersen, B. A.; Schmidt, A. W.; Spracklin, D. K.; Mente, S.; Butler, T. W.;

Howard, H.; Lettiere, D. J.; Rubitski, D. M.; Wong, D. F.; Nedza, F. M.; Nelson, F. R.; Rollema, H.; Raggon, J. W.; Aubrecht, J.; Freeman, J. K.; Marcek, J. M.; Cianfrogna, J.; Cook, K. W.; James, L. C.; Chatman, L. A.; Iredale, P. A.; Banker, M. J.; Homiski, M. L.; Munzner, J. B.; Chandrasekaran, R. Y. Discovery of Two Clinical Histamine H(3) Receptor Antagonists: Trans-N-Ethyl-3-Fluoro-3-[3-Fluoro-4-(Pyrrolidinylmethyl)Phenyl]Cyclobutanecarboxamide (PF-03654746) and Trans-3-Fluoro-3-[3-Fluoro-4-(Pyrrolidin-1-Ylmethyl)Phenyl]-N-(2-Methylpropyl)Cyclobuta. *J. Med. Chem.* **2011**, *54* (21), 7602–7620.

(10)   Sun, M.; Zhao, C.; Gfesser, G. A.; Thiffault, C.; Miller, T. R.; Marsh, K.; Wetter, J.; Curtis, M.; Faghih, R.; Esbenshade, T. A.; Hancock, A. A.; Cowart, M. Synthesis and SAR of 5-Amino- and 5-(Aminomethyl)Benzofuran Histamine H3 Receptor Antagonists with Improved Potency. *J. Med. Chem.* **2005**, *48* (20), 6482–6490.

(11)   Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* **2016**, *54*, 5.6.1-5.6.37.

(12)   Berlin, M.; Lee, Y. J.; Boyce, C. W.; Wang, Y.; Aslanian, R.; McCormick, K. D.; Sorota, S.; Williams, S. M.; West, R. E.; Korfmacher, W. Reduction of HERG Inhibitory Activity in the 4-Piperidinyl Urea Series of H3 Antagonists. *Bioorg. Med. Chem. Lett.* **2010**, *20* (7), 2359–2364.

(13)   Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754.

(14)   RDKit: Open-Source Cheminformatics; Http://Www.Rdkit.Org.

(15)   Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner; 2008; pp 319–326.

4.3 PyRod: Tracing Water Molecules in Molecular Dynamics Simulations

The ligand-guided homology modeling workflow for H$_3$R with subsequent 3D pharmacophore virtual screening resulted in the characterization of critical binding site residue orientations and the identification of two novel ligands with nanomolar activity against H$_3$R (see section 4.2). The presented workflow also delivers binding modes that are essential for optimization of ligands against multiple targets. However, the workflow exploits an essential and well characterized ligand-protein interaction for docking, which is not available for MCHR1. Also, the MCHR1 binding pockets contains several highly flexible binding site residues that complicate docking studies which can be heavily affected by small conformational changes. Thus, a novel method termed PyRod was developed that can derive pharmacophore features from MD simulations. The following publication describes the implementation and algorithms of PyRod and presents the results from a retrospective evaluation against five therapeutically relevant drug targets with available atomistic models and ligand data.

Contribution:

Conceptual design (80 %)
Computational experiments (100 %)
Visualization (100 %)
Manuscript preparation (90 %)

PyRod: Tracing Water Molecules in Molecular Dynamics Simulations

David Schaller,[†] Szymon Pach,[†] and Gerhard Wolber*,[†]

[†]Pharmaceutical and Medicinal Chemistry, Freie Universität Berlin, Königin-Luise-Strasse 2+4, 14195 Berlin, Germany

S *Supporting Information*

ABSTRACT: Ligands entering a protein binding pocket essentially compete with water molecules for binding to the protein. Hence, the location and thermodynamic properties of water molecules in protein structures have gained increased attention in the drug design community. Including corresponding data into 3D pharmacophore modeling is essential for efficient high throughput virtual screening. Here, we present PyRod, a free and open-source Python software that allows for visualization of pharmacophoric binding pocket characteristics, identification of hot spots for ligand binding, and subsequent generation of pharmacophore features for virtual screening. The implemented routines analyze the protein environment of water molecules in molecular dynamics (MD) simulations and can differentiate between hydrogen bonded waters as well as waters in a protein environment of hydrophobic, charged, or aromatic atom groups. The gathered information is further processed to generate dynamic molecular interaction fields (dMIFs) for visualization and pharmacophoric features for virtual screening. The described software was applied to 5 therapeutically relevant drug targets, and generated pharmacophores were evaluated using DUD-E benchmarking sets. The best performing pharmacophore was found for the HIV1 protease with an early enrichment factor of 54.6. PyRod adds a new perspective to structure-based screening campaigns by providing easy-to-interpret dMIFs and purely protein-based pharmacophores that are solely based on tracing water molecules in MD simulations. Since structural information about cocrystallized ligands is not needed, screening campaigns can be followed, for which less or no ligand information is available. PyRod is freely available at https://github.com/schallerdavid/pyrod.
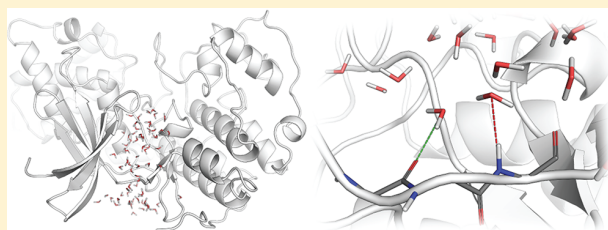
INTRODUCTION

Unliganded protein binding pockets are occupied by water molecules which obligates potential ligands to compete with these water molecules for binding to the protein. Hydrogen bonds between water and protein need to be broken, and replaced water molecules will be released to the bulk solvent. This process heavily affects the thermodynamic properties of the system and renders water as one of the key elements to understand and promote ligand binding.[1,2] Several approaches (e.g., 3D-RISM[3] and GIST[4]) have been developed and employed to estimate the enthalpic and entropic contribution of replacing water molecules from protein binding sites which proved to be useful in pinpointing hot spots for ligand binding and in explaining structure−activity relationships. Including data from molecular dynamics (MD) simulations was found to improve such predictions.[5] However, researchers at GSK conclude in a recent perspective[6] that many studies utilizing water-based methods are of retrospective nature, and several results could have been obtained by simply looking at the atomistic models, e.g., growing a ligand into a hydrophobic protein pocket will most likely increase the affinity.

Pharmacophores describe electrostatic and steric features needed for a molecule to bind to a desired drug target and can be employed in a truly prospective fashion in efficient high-throughput virtual screening campaigns to identify novel active entities.[7] Recently, MD simulations were analyzed to generate so-called water pharmacophores.[8] The researchers analyzed the

thermodynamic characteristics of hydration sites in binding pockets of several drug targets and were able to translate this information into pharmacophores that were successfully evaluated in retrospective screening campaigns. However, the method makes use of commercial software and is not available for public use.

Here, we present PyRod, a free and open-source Python software that was built to translate the highly complex but important information from MD simulations into simplistic and highly efficient pharmacophore models suitable for virtual screening. PyRod supports the 4 major force fields CHARMM,[9] AMBER,[10] GROMOS,[11] and OPLS[12] granting maximum flexibility for the user in choosing the simulation package for generating MD simulation data. We applied PyRod to 5 important drug targets and evaluated its capability to generate successful pharmacophores for virtual screening.

IMPLEMENTATION

PyRod is available as open-source software written in Python 3[13] and employs the external packages MDAnalysis,[14] NumPy,[15] and SciPy.[16] It is composed of several components that can be executed individually via self-explainable config files (Figure 1A). Additionally, a trajectory pharmacophore combo
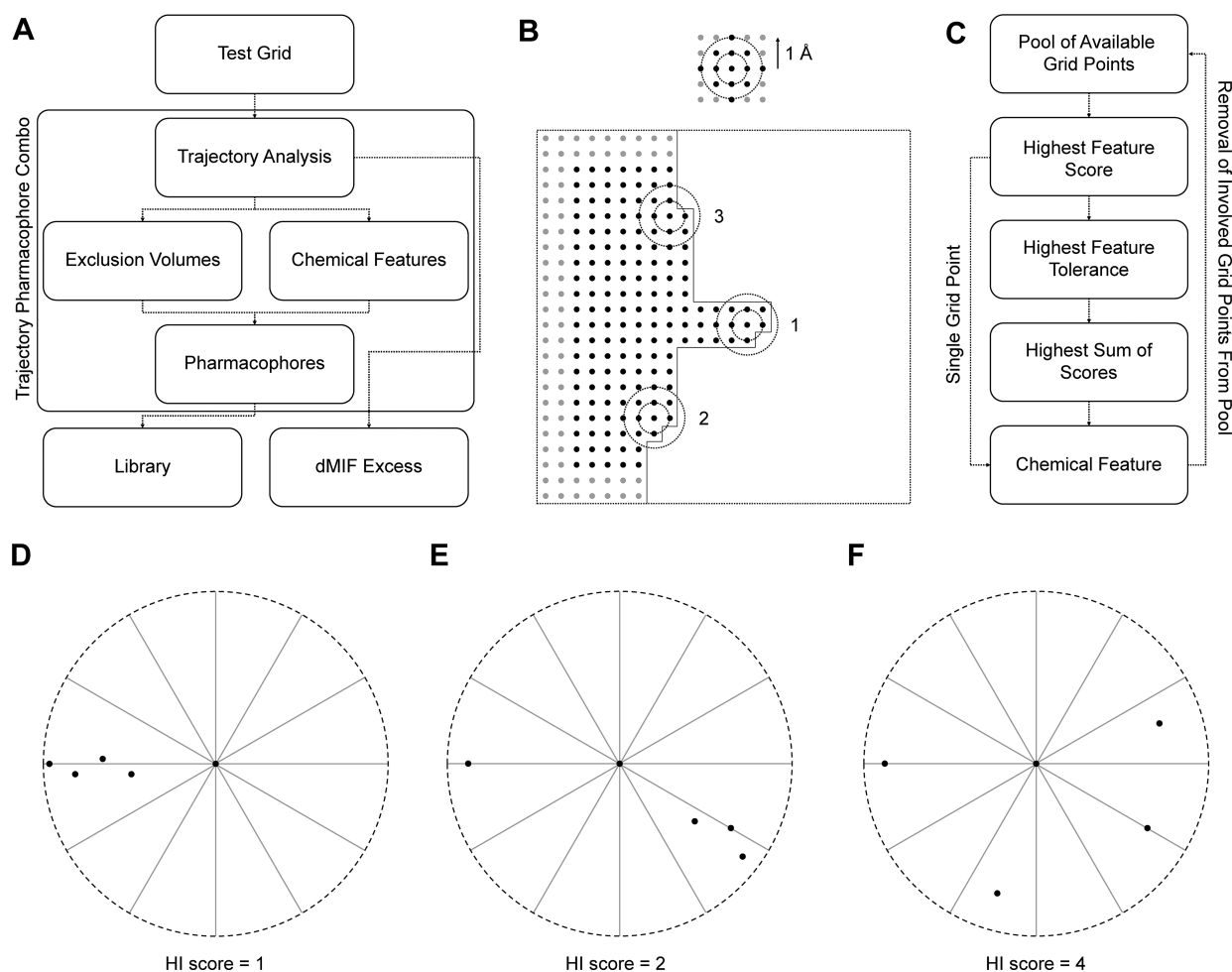
Figure 1. (A) Workflow diagram of PyRod. Each box represents a single component of PyRod that can be executed separately or in combination using the trajectory pharmacophore combo. (B) Depiction of the exclusion volume sphere generation algorithm. Top: The neighborhood within 0.5 and 1 Å of each grid point is analyzed. Bottom: Only grid points with a shape score less than 1 are depicted, representing an area with low water occupancy, e.g., within the protein. Grid points with a shape score of 1 or higher are not depicted and would fall into the white area corresponding to a potential protein binding pocket. Gray grid points are at the grid boundaries and are not considered for exclusion volume placement. The described algorithm favors exclusion volume spheres with a maximum number of neighbors within 0.5 Å, but a minimal number of neighbors within 1 Å and would generate 3 exclusion volumes in this example in the order described by the numbers. (C) Depiction of the iterative chemical feature generation. (D−F) Examples for hydrophobic interaction (HI) scores scaled by buriedness. Each depiction represents a water oxygen with 4 hydrophobic atoms within 5 Å resulting in different HI scores. The buriedness algorithm evaluates the hydrophobic atom positions by calculating angles with the water oxygen as vertex. (D) Water molecule is not buried. (E) Water molecule is buried between two hydrophobic centers. (F) Water molecule is buried between 4 hydrophobic centers.

config file is provided which enables a one-step-execution of several tasks.

Test Grid. This component facilitates the identification of parameters for proper grid placement allowing the user to focus on the protein area of interest in later trajectory analysis. The x, y, and z center parameters are used to define the center of the grid which can be retrieved from, e.g., a central atom in the binding pocket with coordinates stored in the topology file or by employing pocket detection algorithms externally. The x, y, and z edge length parameters define the size of the grid and are usually set between 20 and 30 Å but can be set higher if the whole protein surface should be explored. The test grid will be saved in pdb format with grid points as pseudoatoms. This file can then be visualized together with the topology file to improve parameters if required.

Trajectory Analysis. The implemented routines analyze the protein environment of water molecules in MD simulations to predict favorable sites for chemical feature placement, i.e., hydrogen bond, ionizable, hydrophobic, and aromatic interactions. The employed heuristic scoring functions do not calculate thermodynamic properties but instead estimate favorable regions for chemical feature placement in each frame based on fast-to-calculate geometric descriptors. Since scoring is only performed in the presence of water, PyRod favors regions with stable water molecules whose replacement by ligands will result in a gain of entropy. Chemical feature scores are summated throughout the trajectory in a NumPy[15] representation of the 3D grid whose position and size can be determined by using the test grid component. Spacing between grid points is fixed to 0.5 Å. MDAnalysis[14] is employed as a topology and trajectory reader supporting various molecular

data formats. Residue and atom names are standardized enabling the support of widely used force fields CHARMM,[9] AMBER,[10] GROMOS,[11] and OPLS.[12] The gathered information is transformed for each chemical feature type into dynamic molecular interaction fields (dMIFs) and can be saved in density map format (kont, cns, xplor) for visualization of pharmacophoric binding pocket characteristics. The name dMIF was chosen, since the generated maps introduce dynamics to the concept of molecular interaction fields (MIFs), an established tool in modern drug discovery describing the interaction energy between a probe and a molecular target. For an extensive review on molecular interaction fields we would like to refer the reader's attention to a publication by Artese and coauthors.[17]

In a first step, the protein is analyzed for atoms and atom groups corresponding to potential chemical feature interaction partners (Supporting Information Table S1), e.g., oxygen atoms of the aspartate carboxylate group are hydrogen bond acceptors and part of a negatively charged group if deprotonated. Next, water molecules are localized in each trajectory frame, and their protein environment is analyzed for the previously defined chemical feature interaction partners. If certain geometrical criteria are met (Supporting Information Table S2), grid points within 1.41 Å (radius of water molecule[18]) of the water molecule are identified using fast KDTree[19] routines from SciPy[16] and scored according to the chemical feature type.

Water molecules have two lone pairs and two hydrogens allowing the formation of hydrogen bond as acceptor as well as donor. If water molecules at a certain position are half of the time hydrogen bonded to the protein as donor and half of the time as donor and acceptor, this could indicate two different protein conformations which is important to pharmacophore generation. Hence, hydrogen bonding interactions are split into 6 categories, i.e., single hydrogen bond donor (HD), single hydrogen bond acceptor (HA), double hydrogen bond donor (HD2), double hydrogen bond acceptor (HA2), and mixed hydrogen bond donor/acceptor (HDA). Water molecules involved in more than two hydrogen bonds with the protein are treated as trapped water (TW). Such water molecules are typically deeply buried in the protein making them barely accessible for ligands, and only a very few ligands would be able to fulfill the geometric criteria to replace the water molecule sufficiently. Thus, trapped waters are not considered for later chemical feature generation. However, trapped waters can be of interest in later screening hit selection, since they might serve as a bridge between protein and ligand. Water molecules near metal ions (e.g., Zn^{2+}, Mn^{2+}) are treated as hydrogen bond acceptors and are included in the hydrogen bond count to identify trapped waters. Positions of the protein interaction partners are stored to allow later chemical feature generation with directionality. Gathered hydrogen bond scores are transformed into easy-to-interpret occupancies, e.g., an HA score of 15 means that in 15% of the frames there was a water molecule at this position being involved as a single hydrogen bond acceptor with the protein.

Positive (PI) and negative ionizable interactions (NI) are also scored as occupancies but are additionally scaled by distance, since they represent long-range interactions whose energy decays with increasing distance. Furthermore, PI and NI quench each other. A PI score of 80 can describe very different situations, e.g., there was a water molecule in 80% of the frames in the optimal distance to 1 PI partner but no NI

partner or in 40% of the frames in the optimal distance to 3 PI partners and 1 NI partner.

Aromatic interactions (AI) show a rather complex geometry involving several angles and distances (Supporting Information Table S2). Additionally, in contrast to hydrogen bond and ionizable interactions, water molecules will not necessarily accumulate at a favorable position for potential ligand interaction partners. Thus, grid points close to such water molecule are evaluated individually to satisfy the AI geometry criteria and receive an individually distance scaled occupancy score. Idealized positions for potential interaction partners are stored for later feature generation with directionality.

Cation-π interactions are also recorded by the implemented routines and included in PI and AI scores, e.g., water molecules close to the aromatic ring of a phenylalanine will be scored for PI and water molecules close to the positively charged amine of a lysine will be scored for AI. However, they receive a dedicated heuristic scoring function differing from earlier presented PI and AI scoring (Supporting Information Table S2).

Regions for potential hydrophobic interactions (HI) are identified by counting hydrophobic atoms in the vicinity of each water molecule. This crude atom count is additionally scaled by buriedness to highlight regions with water molecules deeply enclosed in a hydrophobic pocket (Figure 1D−F). In a first step, positions of hydrophobic atoms within 5 Å of the water oxygen are collected. If only one hydrophobic atom was identified, the hydrophobic score is 1, and no further calculation is performed. Otherwise, hydrophobic atom positions are analyzed to estimate the buriedness of the water molecule as follows. Two hydrophobic atom positions are determined that form the maximal angle with the water oxygen position as vertex, e.g., two hydrophobic atoms with a water molecule exactly in between would lead to an angle of 180 degrees. If this maximal angle is less than 30 degrees, the algorithm is terminated, and the hydrophobic score remains 1 (Figure 1D). Such a situation corresponds to a geometry, where a water molecule is close to multiple hydrophobic atoms but not buried. Otherwise, the hydrophobic score is increased by 1, both hydrophobic atom positions are marked as accepted, and one of the two hydrophobic atom positions forming the maximal angle is randomly selected as the reference position for further processing. The remaining hydrophobic atom positions are analyzed in an iterative fashion as follows. First, the hydrophobic atom position is identified that forms the maximal angle with the reference position and the water oxygen position as vertex. Next, this hydrophobic atom position is evaluated for angles formed with the already accepted positions and the water oxygen as vertex. If none of the formed angle is smaller than 30 degrees, the hydrophobic score is increased by 1, and the evaluated position is marked as accepted position. Otherwise, this position is ignored. This procedure is repeated until all hydrophobic atom positions were evaluated resulting in a hydrophobic score that is strongly dependent on the hydrophobic buriedness of the water molecule. However, highly hydrophobic regions may not be sampled well by water molecules and will be scored less frequently and consequently may receive a lower absolute HI score. Hence, a normalized score is provided as well (HI_{norm}) which reports the average hydrophobic score per occurrence, e.g., an HI_{norm} score of 5.3 means that when a water molecule occurred at this position the near grid points retrieved on average an HI score of 5.3.

Exclusion Volume Spheres. During trajectory analysis the presence of water is recorded to generate a shape dMIF. It is used in this component to place exclusion volumes limiting the binding pocket volume in later pharmacophores with the following algorithm (Figure 1B). Grid points must have a shape score less than 1 (corresponding to water occupancy of 1%, can be changed by the user) and are sorted for the number of neighbors (grid points with a shape score less than 1) within 1 Å. Next, each grid point is evaluated as center of an exclusion volume starting with the grid point with the lowest number of neighbors within 1 Å. Grid points with very few neighbors usually correspond to protein side chains pointing inside the binding pocket and are prioritized by the algorithm. To be accepted as the center of an exclusion volume, a grid point must not be at the grid boundaries, must have exactly 7 neighbors within 0.5 Å but less than 33 neighbors within 1 Å ensuring that exclusion volumes are placed only at the interface of protein and water but not too close to the chemical features, and finally, must not be within 4 Å (2 Å if restrictive mode is enabled) of an already generated exclusion volume.

Chemical Features. A novel algorithm was implemented translating dMIFs into corresponding chemical features for pharmacophore virtual screening (Figure 1C). First, all grid points become part of a pool of available grid points for the respective chemical feature generation. Next, the grid points with the highest feature score in the grid point pool are determined. If this search results in a single grid point, its position will be used as center of the chemical feature. The tolerance radius of that chemical feature is identified by iteratively increasing the search radius (minimum = 1.5 Å) from the feature center in 0.5 Å steps. If the feature score of a grid point within the search radius is below half of the current highest feature score, the search is stopped, and the current search radius will be used as a tolerance radius for the chemical feature. If multiple grid points share the highest feature score in the grid point pool, the following procedures are performed to select a single grid point as the feature center. Tolerance radii are calculated for each of the considered grid points. The grid point with the highest tolerance radius will be used as center of the chemical feature. If multiple grid points share the highest tolerance radius, the sum of feature scores of the grid points within the tolerance radius is calculated, and the grid point with the highest sum of feature scores is selected as the feature center. If this procedure does not lead to the selection of a single grid point, a random pick of the remaining grid points is performed. Grid points within the tolerance radius of a chemical feature must not be part of an already generated chemical feature of the respective chemical feature type. This criterion prevents overlap of chemical features within chemical feature types. In the case of hydrogen bond and aromatic interactions, recorded positions of interaction partners are clustered by searching for the interaction partner position with the most neighboring interaction partner positions within 1.5 Å. This procedure allows the generation of chemical features with directionality. Finally, grid points within the tolerance radius of the feature center are removed from the pool of available grid points, and a new iteration is started. The chemical feature generation is terminated if 20 chemical features (can be changed by the user) of the respective chemical feature type were generated or if the highest feature score of the grid point pool decreases below 1.

Pharmacophores. The output of exclusion volume and chemical feature generation is merged and saved as a single "super pharmacophore" containing all previously generated chemical features and exclusion volumes. Additionally, a pharmacophore can be saved containing only the highest ranked features for each chemical feature class as specified by the user, e.g., 20 highest scored hydrogen bonding features and 5 highest scored hydrophobic features. Currently, PyRod supports LigandScout[20] and pdb-like pharmacophore formats. The pdb-like pharmacophore file uses the residue name column to specify the chemical feature type and aims at providing a pharmacophore format readable by human and various molecular modeling softwares. However, directionality is not included in the pdb-like pharmacophore format.

Combinatorial Library. The generated "super pharmacophores" can contain more than 100 chemical features, which remains computationally challenging to screen also with the current progress in CPU performance. Thus, reducing the number of chemical features is key to enable fast high-throughput virtual screening. This component facilitates the generation of a combinatorial library of pharmacophores with a specified number of chemical features as defined by the user. First, the user preselects chemical features of interest in LigandScout[20] and saves this pharmacophore for combinatorial processing. Chemical features that should be present in every generated pharmacophore have to be set mandatory, whereas chemical features that should be added in a combinatorial fashion have to be set optional. Next, the user can specify the limits for minimal and maximal number of chemical features in the config file, i.e., number of independent chemical features, number of hydrogen bonding features, number of ionizable features, number of aromatic features, and number of hydrophobic features. Prior to library generation the user will be informed about the number of possible pharmacophores and prompted for execution. To further limit the library size, each pharmacophore is evaluated for the following rules, i.e., (i) ionizable and hydrophobic features should not appear within 3 Å, (ii) different hydrogen bonding features should not be present within 1.5 Å, since such a situation implies two different protein conformations, (iii) different ionizable features should not be present within 3 Å, and (iv) hydrogen bonding features of HA2, HD2, and HDA are not allowed to be split. PyRod also provides a customizable pharmacophore evaluation script written in Python performing receiver operatic characteristics analysis with LigandScout.[20]

dMIF Excess. Selectivity within a protein family as well as the occurrence of mutation-induced resistance remains a major challenge in modern drug discovery.[21,22] It would be desirable to exploit such minor differences in protein binding pockets. Thus, this component enables the comparison of dMIFs between closely related proteins by generating dMIF excess maps visualizing the excess of one system over the other.

■ TEST SYSTEMS

PyRod performance was evaluated on 5 important drug target test systems, i.e., cyclin-dependent kinase 2-cyclin A complex (CDK2, 5if1[23]), HIV-1 protease (HIV1P, 1nh0[24]), estrogen receptor alpha (ERα, 1xpc[25]), dopamine D3 receptor (D3R, 3pbl[26]), and adenosine A_{2A} receptor (A_{2A}R, 5iu4[27]). Protein selection was based on therapeutic relevance, availability of benchmarking sets from DUD-E,[28] and crystal structures from PDB[29] as well as protein family diversity.

System Setup. Crystal structures were retrieved from PDB[29] and prepared in MOE 2015[30] as follows. Ligands were deleted as well as water more than 5 Å away from the protein.

Figure 2. (A) Shape dMIF of CDK2 (cutoff 1) with exclusion volumes present in each generated pharmacophore. (B, C) Characterized binding pocket with dMIFs for single hydrogen bond donor (green, cutoff 38), single hydrogen bond acceptor (red, cutoff 17), mixed hydrogen bond donor/acceptor (cyan, cutoff 14), positive ionizable (blue, cutoff 27), aromatic interaction (magenta, cutoff 36), and hydrophobic interaction (yellow, cutoff 100). Cutoffs were chosen to visualize decision making in subsequent pharmacophore feature selection. (D) Selected pharmacophore features based on chemical feature score and arrangement (green arrow—hydrogen bond donor, red arrow—hydrogen bond acceptor, yellow sphere—hydrophobic interaction, blue star—positive ionizable, blue ring plane—aromatic interaction). (E) Performance evaluation of pharmacophore library. (F) Pharmacophore with best early enrichment factor ($EF_{1;5;10;100}$: 30.3;30.3;30.3;30.3, $AUC_{1;5;10;100}$: 0.99;0.99;0.88;0.54).

Errors were corrected with the Structure Preparation tool. The low resolution D3R structure 3pbl misses a sodium ion that is known to be crucial for inactive class A GPCR states.[31] Hence the sodium ion and 6 coordinating water molecules were transferred from the high resolution structure of δ opioid receptor (4n6h[32]) into the D3R system. Chain breaks were capped with ACE and NME. Protonation states were assigned using the Protonate 3D tool at pH 7. Nonmembrane proteins (CDK2, HIV1P, and ERα) were solvated in a cubic box with TIP4P water, 0.15 M NaCl, and 10 Å padding using Maestro 11.3.[33] Membrane proteins (A$_{2A}$R and D3R) were embedded in a POPC bilayer according to the orientation provided by the PPM server[34] and solvated in an orthorhombic box with TIP4P water, 0.15 M NaCl, and 10 Å padding.

Molecular Dynamics Simulation. Simulations were performed with Desmond 5.1[35] and the OPLS 2005[36] force field on a Nvidia GeForce GTX 1070 graphics card. Minimization and equilibration were performed with default settings. Ten replications of 10 ns simulation were performed for each system with periodic boundary conditions in an NPT ensemble. The temperature was retained at 300 K using the Nose-Hoover thermostat, and the pressure was maintained at 1.01325 bar using the Martyna-Tobias-Klein barostat. Coordinates were saved every 5 ps resulting in 2000 frames per simulation. Trajectories were additionally processed in VMD 1.9.2,[37] i.e., the protein was centered in the water box by using the pbc tool and trajectories were aligned on the protein backbone heavy atoms using the RMSD Trajectory tool.

The CDK2 system (5if1[23]) was also simulated with OpenMM 7.2.2[38] on a Nvidia GeForce GTX 1070 graphics card employing the Amber force field ff14SB[10] with the TIP4P-Ew water model to test the effect of restraining heavy atoms. The same prepared protein structure was used as for Desmond simulations described above. The protein was solvated in a cubic water box with 10 Å padding and 0.15 M NaCl. The Particle Mesh Ewald method was used to calculate long-range electrostatic interactions with a 10 Å cutoff, and all bonds involving hydrogens were constrained in length. Langevin dynamics was performed at 300 K with 2 fs time step. Ten replications of 10 ns simulations were performed with periodic boundary conditions and an NPT ensemble. Coordinates were saved every 5 ps resulting in 2000 frames per simulation. The simulations were performed with and without a custom force of 5 kcal restraining protein heavy atoms at their initial position. Resulting trajectories were processed in VMD[37] as already described.

PyRod. Grid parameters were adjusted using the test grid component to center the grid in the binding site. Grids were cubic with edge lengths of 20 Å for ERα and CDK2, 25 Å for D3R and HIV1P, and 30 Å for A$_{2A}$R. Trajectories were processed with PyRod 0.7.1 using the last 5 ns of each replication resulting in 10,000 frames for analysis of each system with default settings. Generated dMIFs were visualized and analyzed in LigandScout 4.2[20] to preselect pharmacophore features according to feature scores and their arrangement in the binding pocket. Selected chemical features were subjected to combinatorial processing with the combinatorial library component of PyRod. Chemical feature limits for each target can be found in the Supporting Information (Table S3).

Pharmacophore Screening. The ligand benchmarking sets with actives and decoys for CDK2, HIV1P, ERα, D3R, and A$_{2A}$R were retrieved from the DUD-E server[28] in SMILES format. For CDK2, ERα, D3R, and A$_{2A}$R 25 conformations

were generated per molecule with iCon as implemented in LigandScout 4.2.[20] For HIV1P 200 conformations were generated per molecule, since the active set primarily contains peptidomimetics with many rotatable bonds. These databases were used for pharmacophore evaluation in LigandScout 4.2[20] employing receiver operating characteristic (ROC) curves.

■ RESULTS AND DISCUSSION

CDK2. The ATP binding pocket of CDK2 is a well characterized site for inhibition with a plethora of crystal structures deposited in the PDB. The most frequently observed interactions include hydrogen bonds formed with the backbone of residues E81 and L83.[39] Concordantly, PyRod identified a hydration site at which water molecules act as a single hydrogen bond donor to the backbone oxygen of E81 in 63% of all frames (Figure 2B). Besides being involved in a single hydrogen bond, these water molecules are also in a very hydrophobic environment (HI score = 420, HI$_{norm}$ score = 5.85). These characteristics render this hydration site as an essential position for ligand binding, since replacing restrained water molecules from a hydrophobic pocket by a corresponding ligand moiety should be beneficial for the entropy and enthalpy of the system. Adjacent to E81 are further hydration sites at which water molecules are involved in hydrogen bonds with the backbone of L83 in 40% of the frames as single donor, in 19% of the frames as single acceptor, and in 15% of the frames as mixed donor and acceptor (Figure 2B). Likewise, these hydration sites lie in a hydrophobic environment (HI score = 180−350, HI$_{norm}$ score = 3.5−6.25) highlighting these positions for additional ligand interactions. Several hotspots were identified for placing positive ionizable groups, i.e., at the interface of D145 and F80 as well as next to D86 with PI scores of 70 and 25, respectively, and aromatic moieties, i.e., close to the salt bridge formed by K33 and D145 as well as adjacent to F80 with AI scores of 45 and 70, respectively (Figure 2C). A hydrophobic band with HI scores ranging from 100 to 300 is spanning the binding pocket that resulted in the generation of 6 hydrophobic features (Figure 2C). In total 15 chemical features were selected based on the corresponding feature score and their arrangement in the binding pocket (Figure 2D). By employing the combinatorial library component of PyRod, these features were combined to 816 pharmacophores with 3 to 5 chemical features. The hydrogen bonding donor feature pointing toward the backbone oxygen of E81 was selected to be present in every pharmacophore. Further parameters can be found in the Supporting Information (Table S3). Using LigandScout[20] all pharmacophores were screened against an active set retrieved from the DUD-E[28] database and additionally evaluated against decoys if 5% of the actives were found. Hit lists were evaluated for early enrichment factor (EF$_{1\%}$) and plotted against found actives to select pharmacophores of interest (Figure 2E). The most selective pharmacophore (EF$_{1\%}$ = 30.3) consists of 2 hydrophobic features, 1 hydrogen bond donor, and 1 hydrogen bond mixed donor/acceptor (Figure 2F).

D3R. Currently, only a single crystal structure of D3R is available. However, this target was studied extensively with many known ligands stored in public databases. A key interaction shared across all aminergic GPCRs is a charged interaction to an aspartate in the orthosteric binding pocket.[40] PyRod located several hydration sites with water molecules pointing as a single hydrogen bond donor toward D110 in 35−50% of the frames with PI scores between 35 and 50 (Figure
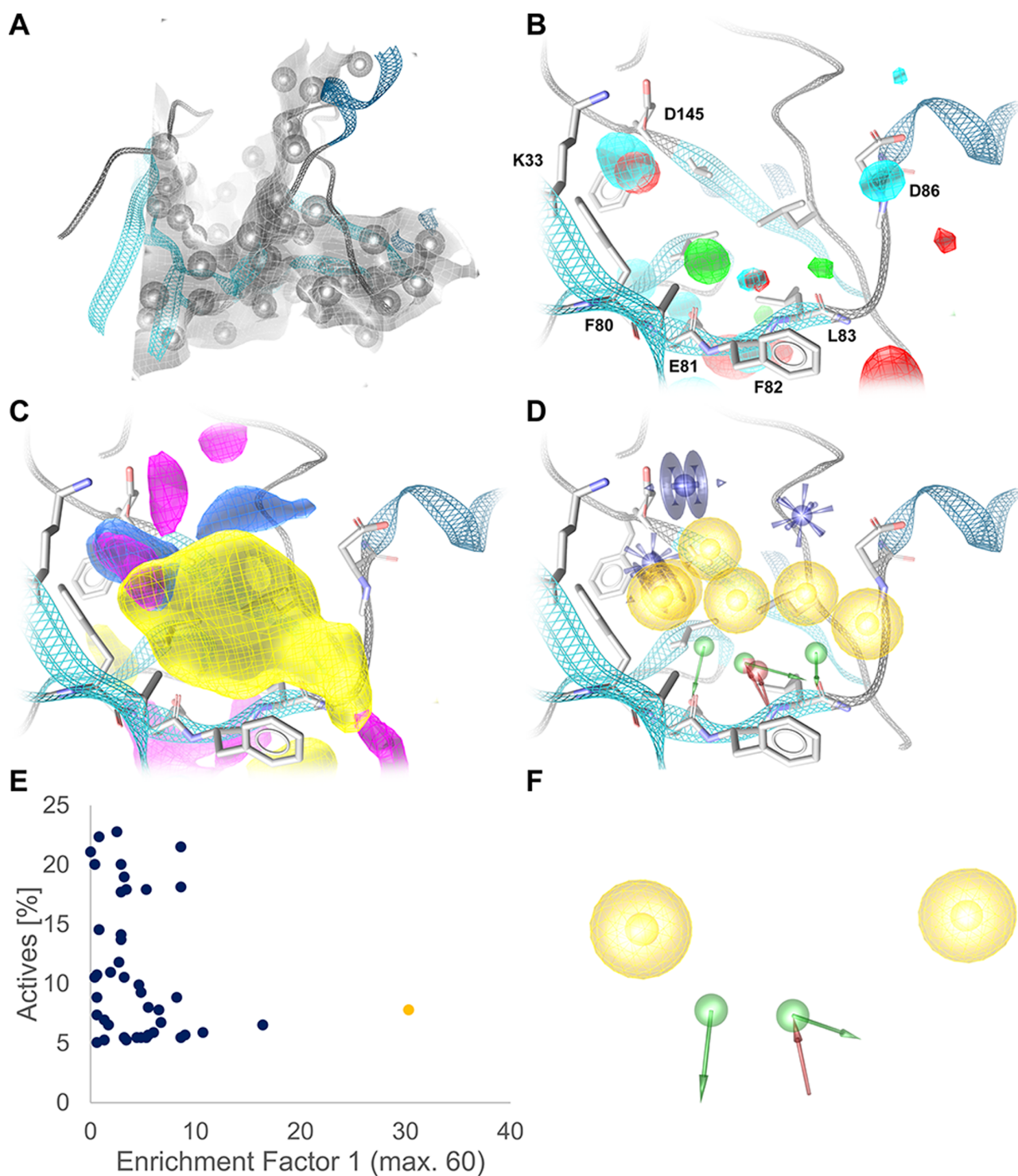
Figure 3. (A) Shape dMIF of D3R (cutoff 1) with exclusion volumes present in each generated pharmacophore. (B, C) Characterized binding pocket with dMIFs for single hydrogen bond donor (green, cutoff 36), single hydrogen bond acceptor (red, cutoff 19), positive ionizable (blue, cutoff 33), aromatic interaction (magenta, cutoff 14), and hydrophobic interaction (yellow, cutoff 160). Cutoffs were chosen to visualize decision making in subsequent pharmacophore feature selection. (D) Selected pharmacophore features based on chemical feature score and arrangement (green arrow—hydrogen bond donor, red arrow—hydrogen bond acceptor, yellow sphere—hydrophobic interaction, blue star—positive ionizable, blue ring plane—aromatic interaction). (E) Performance evaluation of pharmacophore library. (F) Pharmacophore with best early enrichment factor ($EF_{1;5;10;100}$: 7.7;3.3;2.3;2.0, $AUC_{1;5;10;100}$: 0.92;0.94;0.93;0.58).

3B,C). Additionally, we found 2 sites with water molecules acting as hydrogen bond acceptors with an HA score of 20 next to the backbone nitrogen of I183 and to the imidazole ring of H349. Hydrophobic hotspots (HI score = 140−225, HI_{norm}
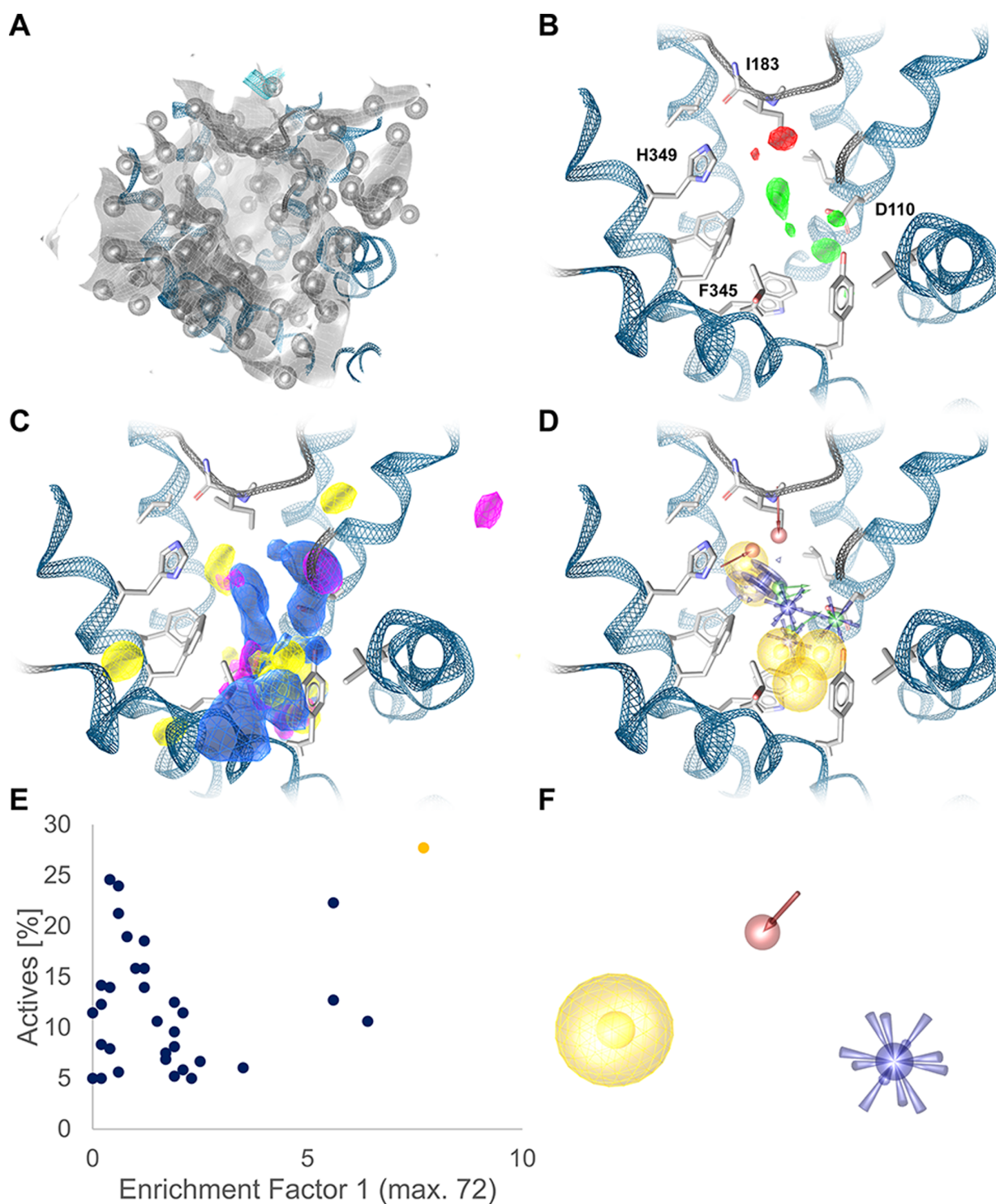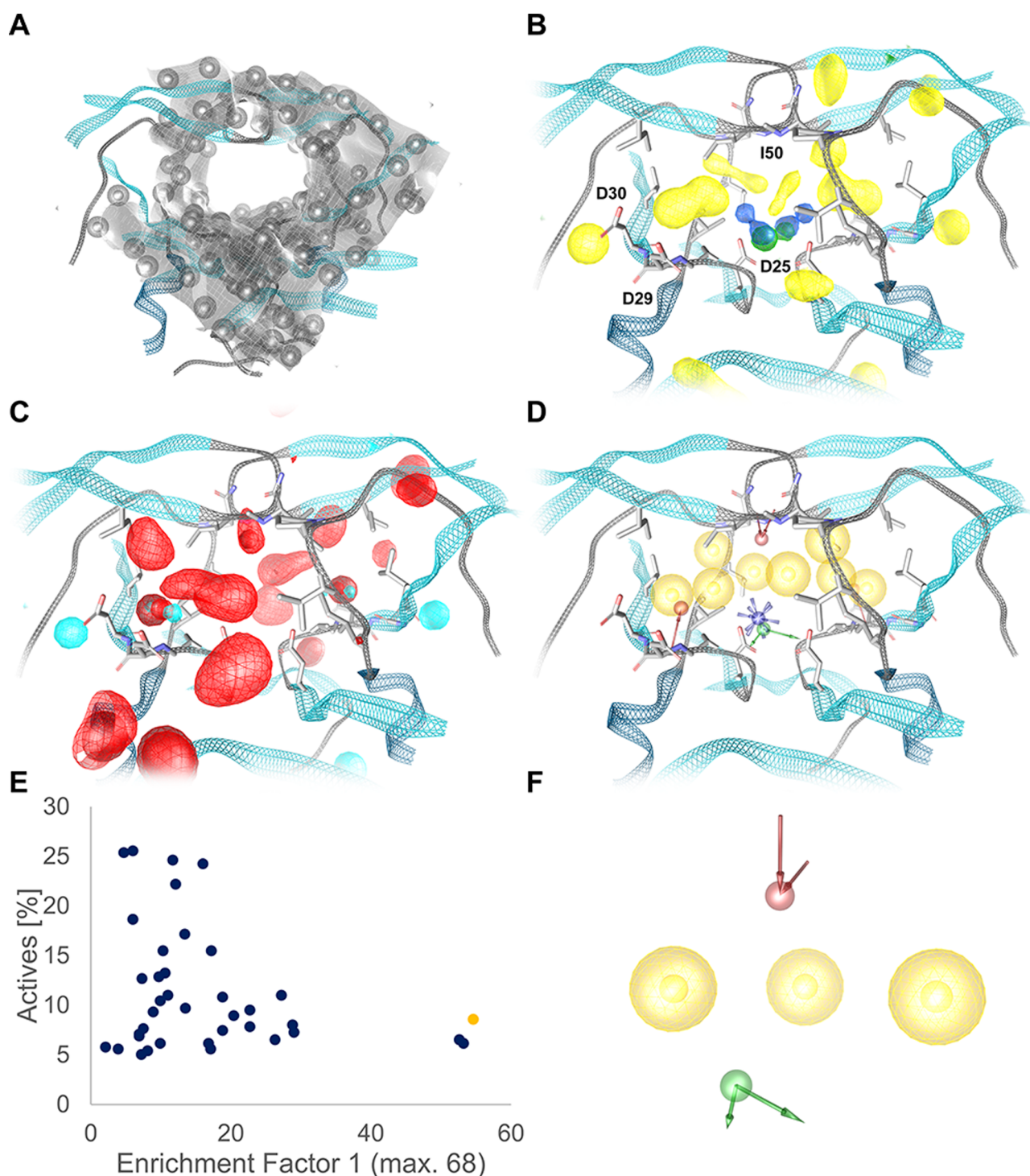
Figure 4. (A) Shape dMIF of HIV1P (cutoff 1) with exclusion volumes present in each generated pharmacophore. (B, C) Characterized binding pocket with dMIFs for double hydrogen bond donor (dark green, cutoff 35), positive ionizable (blue, cutoff 90), hydrophobic interaction (yellow, cutoff 110), single hydrogen bond acceptor (red, cutoff 13), double hydrogen bond acceptor (dark red, cutoff 4), and mixed hydrogen bond donor/acceptor (cyan, cutoff 26). Cutoffs were chosen to visualize decision making in subsequent pharmacophore feature selection. (D) Selected pharmacophore features based on chemical feature score and arrangement (green arrow—hydrogen bond donor, red arrow—hydrogen bond acceptor, yellow sphere—hydrophobic interaction, blue star—positive ionizable). (E) Performance evaluation of pharmacophore library. (F) Pharmacophore with best early enrichment factor ($EF_{1;5;10;100}$: 54.6;54.6;54.6;54.6, $AUC_{1;5;10;100}$: 1.00;1.00;0.93;0.54).

score = 3.80−4.25) were identified close to F345 and above D110 as well as sites for aromatic interactions (AI score = 15) next to F345. In total, 16 chemical features were selected and

combined to 2441 pharmacophores with 3 to 5 independent features (Figure 3D). Further parameters can be found in the Supporting Information (Table S3). The best performing
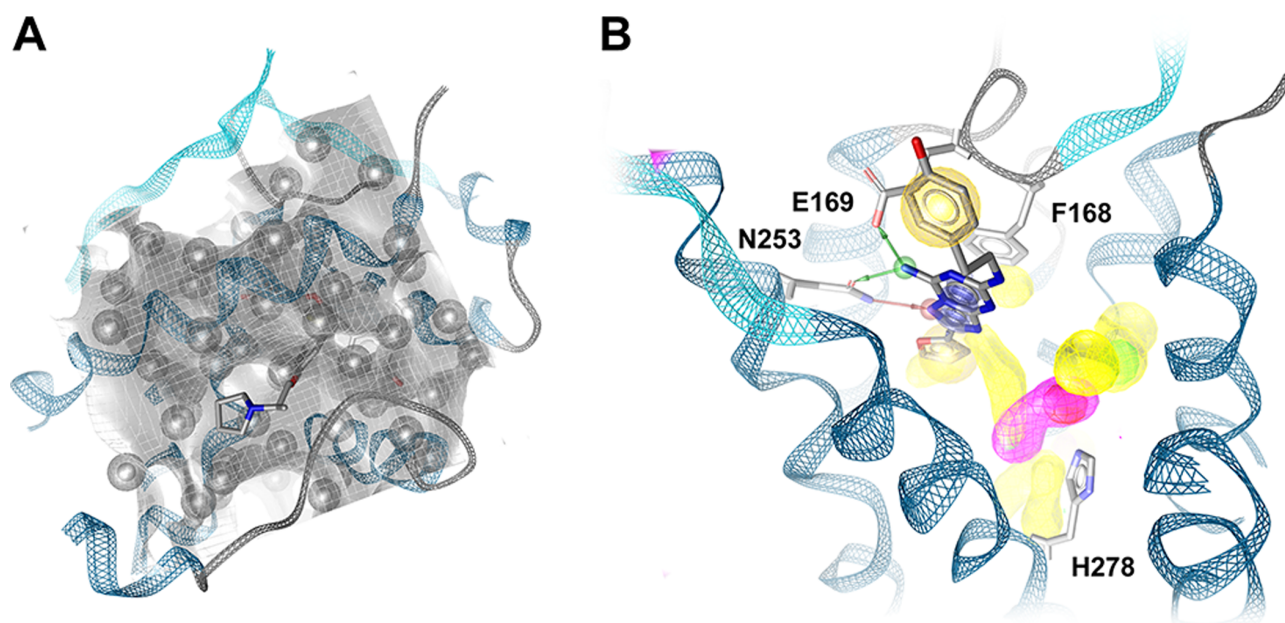
Figure 5. (A) Shape dMIF of ERα (cutoff 1) with an exclusion volume placed at a position where cocrystallized ligands bind. (B) Characterized binding pocket of A_2AR with dMIFs for hydrogen bond donor (green, cutoff 40), single hydrogen bond acceptor (red, cutoff 40), aromatic interaction (magenta, cutoff 32), and hydrophobic interaction (yellow, cutoff 160). Essential key interactions known from cocrystallized ligands are not represented in the dMIFs (green arrow—hydrogen bond donor, red arrow—hydrogen bond acceptor, blue ring plane—aromatic interaction, yellow sphere—hydrophobic interaction).

pharmacophore ($EF_{1\%}$ = 7.7) consists of 1 hydrophobic feature, 1 hydrogen bond acceptor, and 1 positive ionizable feature (Figure 3E,F).

**HIV1P.** The protease of HIV1 is a well characterized target for inhibiting virus replication. Mature HIV1P exists as a homodimer with two aspartates (D25, D25′) in the catalytic center. The hydroxyl group of approved drugs mimics a water molecule present in the transition state and results in inhibition of the protease.[41] PyRod located two hydration sites between the two catalytic aspartates with HD2 scores of 40 and a PI score of 110 (Figure 4B). Hydrophobic pockets are symmetrically distributed around the catalytic center resulting in several hydrophobic features (HI score = 120−270, HI_{norm} score = 2.80−4.90). PyRod also found the hydration site (Figure 4C) between the backbone nitrogens of I50 and I50′ where water molecules are bound as a double or single hydrogen bond acceptor (HA2 score = 5, HA score = 17). This water molecule is observed frequently in inhibitor-bound crystal structures serving as a bridge between ligand and protein but can also be replaced.[41] Finally, PyRod identified a hydration site next to D29 and D30 with water molecules acting as double hydrogen bond acceptor (HA2 score = 6), single hydrogen bond acceptor (HA score = 15), or mixed hydrogen bond donor/acceptor (HDA score = 27). A single hydrogen bond acceptor feature was chosen at this position to represent all 3 observed water conformations (Figure 4D). In total, 588 pharmacophores were generated by combining 15 pharmacophore features. The double hydrogen bond donor feature was selected to be present in every pharmacophore. Further parameters can be found in the Supporting Information (Table S3). The best performing pharmacophore ($EF_{1\%}$ = 54.6) consists of 3 hydrophobic features, 2 hydrogen bond donors, and 2 hydrogen bond acceptors (Figure 4E-F).

ERα and A_2AR. Estrogen receptor alpha and adenosine A_2A receptor represent 2 test cases for which pharmacophore generation based on water dynamics was not successful. ERα contains a hydrophobic pocket[25] that is collapsing upon unrestrained molecular dynamics simulation. This ultimately leads to the placement of exclusion volumes at a position where cocrystallized ligands bind (Figure 5A). Agonists and antagonists of A_2AR share two key interactions with F168 and N253.[42] The aromatic interaction with F168 is completely absent in the AI dMIF, since F168 is very flexible in the unbound state and adapts a conformation differing from the ligand bound state (Figure 5B). Also, N253 and E169 leave the ligand bound conformation quickly upon initiating unrestrained simulations and do not frequently interact with water molecules at positions known from ligand interaction with N253.

Discussion. An important decision to make when using PyRod is the MD simulation length. To estimate the equilibration process of water molecules in protein binding pockets, the change of water occupancy in trajectory bins of 50 frames was analyzed for each test system and plotted together with the RMSD of protein heavy atoms (Supporting Information Figure S1). All test systems in this study were equilibrated within the first 5 ns of unrestrained MD simulation. However, this may be different for other systems. Interestingly, these plots indicate a synchronous equilibration of protein and water rendering the protein RMSD an easy-to-use descriptor to estimate equilibration times of water molecules in protein binding pockets. The total simulation length was restricted to 10 ns for each replication to reduce computational costs and to sample protein conformations close to the crystallographic structure. Replications were performed to expand sampling of protein conformations without
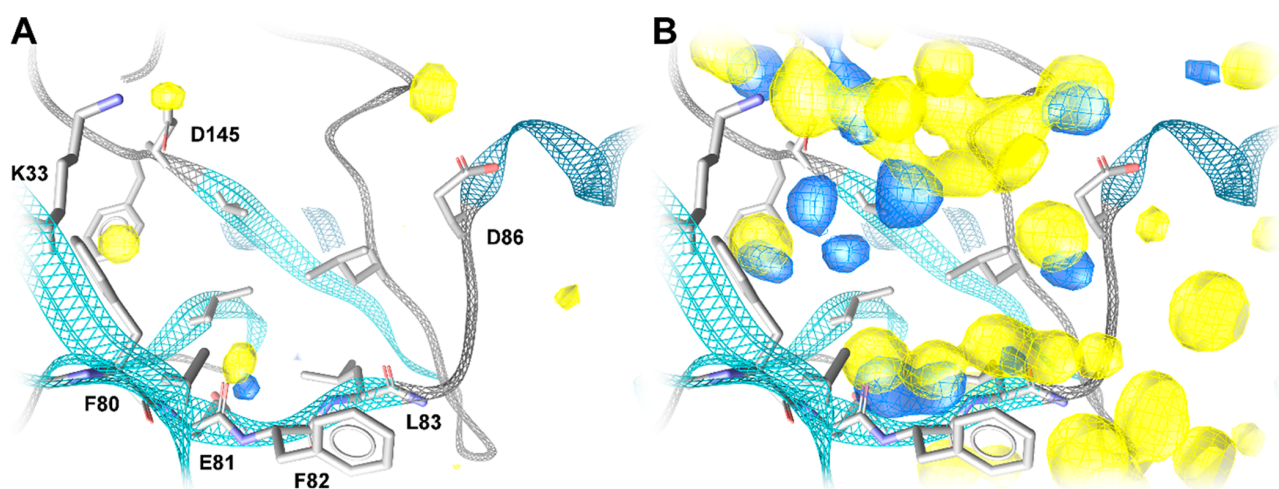
Figure 6. Effect of restraining protein heavy atoms on the binding pocket hydration sites of the CDK2 system. Generated dMIFs for hydrogen bonding interactions (HB) differ significantly between simulations with restrains on protein heavy atoms (yellow) and without (blue). HB dMIFs show areas with a HB score of at least 90% of the maximal HB score (A) and at least 50% of the maximal HB score (B).

introducing artifacts from a single long MD simulation stuck in a local minimum.[43]

Best PyRod pharmacophores of CDK2, D3R, and HIV1P outperform the docking program DOCK 3.6 when comparing early enrichment factors ($EF_{1\%}$) with the DUD-E benchmark[28] (Supporting Information Table S4). However, PyRod pharmacophores could not be generated for $A_{2A}R$ and $ER\alpha$, since both targets quickly leave the ligand-bound conformation upon unrestrained MD simulation. Although restraining the protein heavy atoms is tempting, this procedure would neglect the contribution of the protein to the entropy of the system.[44] When restraining heavy atoms of the CDK2 system, we observed many more stable hydration sites with overall higher feature scores, which may hinder prioritization of important chemical features (Figure 6). Instead of restraining the protein in MD simulations, it might be an option in such situations to employ methods that generate pharmacophores based on the static structure.[45,46]

An important difference between the already published water pharmacophore method[8] and our PyRod approach is the number of generated pharmacophores. The water pharmacophore method was designed to generate a single pharmacophore in a highly automated fashion. Although retrospectively successful with 4 out of 7 targets, it needs to be shown that the parameters and cutoffs trained on the test systems also succeed in a prospective study on a completely different target. PyRod does not generate a single pharmacophore for virtual screening but a combinatorial library. This agrees with the fact that different ligands can show different interaction patterns for the same binding pocket. However, it would be desirable to develop only a few diverse pharmacophore models with PyRod without the knowledge of any ligand data. Prospective studies are forthcoming and are the only possibility to prove the usefulness of PyRod in such situations. Nevertheless, we are confident that we would have been able to generate successful pharmacophores by only analyzing dMIFs and selecting features for CDK2 and HIV1P.

CONCLUSION

In this study we show that water dynamics from MD simulations can be used to generate highly usable 3D

pharmacophore models for virtual screening. Employing the free and open-source software PyRod we were able to successfully describe pharmacophoric binding pocket characteristics and generate pharmacophores for three pharmaceutically relevant drug targets. The early enrichment factors from the best performing models range from 7.7 for D3R to 54.6 for HIV1P. Additionally, we found that restraining protein heavy atoms dramatically affects the water dynamics in the binding pocket hindering hot spot identification for ligand binding in water-based methods.

ASSOCIATED CONTENT

S **Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.9b00281.

> Tables for feature definitions, scoring functions, combinatorial library parameters, and early enrichment factor comparison with DUD-E benchmark (PDF)

AUTHOR INFORMATION

Corresponding Author
*E-mail: gerhard.wolber@fu-berlin.de.
ORCID
David Schaller: 0000-0002-1881-4518
Gerhard Wolber: 0000-0002-5344-0048

Author Contributions

D.S. designed the software. D.S., S.P., and G.W. designed the experiments. D.S. and S.P. performed and analyzed the experiments. G.W. directed the studies.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

ABBREVIATIONS

3D-RISM, 3D reference interaction site model; $A_{2A}R$, adenosine A_{2A} receptor; AI, aromatic interaction; CDK2,

cyclin-dependent kinase 2; D3R, dopamine D3 receptor; dMIF, dynamic molecular interaction field; ERα, estrogen receptor alpha; GIST, grid inhomogeneous solvation theory; HA, single hydrogen bond acceptor; HA2, double hydrogen bond acceptor; HB, hydrogen bond; HD, single hydrogen bond donor; HD2, double hydrogen bond donor; HDA, mixed hydrogen bond donor/acceptor; HI, hydrophobic interaction; HIV1P, human immunodeficiency virus 1 protease; MD simulation, molecular dynamics simulation; NI, negative ionizable; PI, positive ionizable; TW, trapped water

■ REFERENCES

(1) Nittinger, E.; Schneider, N.; Lange, G.; Rarey, M. Evidence of Water Molecules—A Statistical Evaluation of Water Molecules Based on Electron Density. *J. Chem. Inf. Model.* **2015**, *55* (4), 771−783.

(2) Spyrakis, F.; Ahmed, M. H.; Bayden, A. S.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E. The Roles of Water in the Protein Matrix: A Largely Untapped Resource for Drug Discovery. *J. Med. Chem.* **2017**, *60* (16), 6781−6827.

(3) Heil, J.; Kast, S. M. 3D RISM Theory with Fast Reciprocal-Space Electrostatics. *J. Chem. Phys.* **2015**, *142* (11), 114107.

(4) Nguyen, C. N.; Kurtzman Young, T.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]Uril. *J. Chem. Phys.* **2012**, *137* (4), 044101.

(5) Bucher, D.; Stouten, P.; Triballeau, N. Shedding Light on Important Waters for Drug Design: Simulations versus Grid-Based Methods. *J. Chem. Inf. Model.* **2018**, *58* (3), 692−699.

(6) Graves, A. P.; Wall, I. D.; Edge, C. M.; Woolven, J. M.; Cui, G.; Le Gall, A.; Hong, X.; Raha, K.; Manas, E. S. A Perspective on Water Site Prediction Methods for Structure Based Drug Design. *Curr. Top. Med. Chem.* **2017**, *17* (23), 2599−2616.

(7) Seidel, T.; Ibis, G.; Bendix, F.; Wolber, G. Strategies for 3D Pharmacophore-Based Virtual Screening. *Drug Discovery Today: Technol.* **2010**, *7* (4), e221−e228.

(8) Jung, S. W.; Kim, M.; Ramsey, S.; Kurtzman, T.; Cho, A. E. Water Pharmacophore: Designing Ligands Using Molecular Dynamics Simulations with Water. *Sci. Rep.* **2018**, *8* (1), 10400.

(9) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14* (1), 71−73.

(10) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696−3713.

(11) Reif, M. M.; Hünenberger, P. H.; Oostenbrink, C. New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field. *J. Chem. Theory Comput.* **2012**, *8* (10), 3705−3723.

(12) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11* (7), 3499−3509.

(13) van Rossum, G. *Python Reference Manual*; 1995.

(14) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32* (10), 2319−2327.

(15) Oliphant, T. E. *Guide to Numpy*, 2nd ed.; 2015.

(16) Jones, E.; Oliphant, T. E.; Peterson, P. *SciPy: Open Source Scientific Tools for Python*; 2001.

(17) Artese, A.; Cross, S.; Costa, G.; Distinto, S.; Parrotta, L.; Alcaro, S.; Ortuso, F.; Cruciani, G. Molecular Interaction Fields in Drug Discovery: Recent Advances and Future Perspectives. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3* (6), 594−613.

(18) Franks, F. Structure of the Water Molecule and the Nature of the Hydrogen Bond in Water. In *Water*; Franks, F., Ed.; Royal Society of Chemistry: Cambridge, 2000; pp 9−14, DOI: 10.1039/9781847552341-00009.

(19) Bentley, J. L. Multidimensional Binary Search Trees Used for Associative Searching. *Commun. ACM* **1975**, *18* (9), 509−517.

(20) Wolber, G.; Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2005**, *45* (1), 160−169.

(21) Huggins, D. J.; Sherman, W.; Tidor, B. Rational Approaches to Improving Selectivity in Drug Design. *J. Med. Chem.* **2012**, *55* (4), 1424−1444.

(22) Hao, G.-F.; Yang, G.-F.; Zhan, C.-G. Structure-Based Methods for Predicting Target Mutation-Induced Drug Resistance and Rational Drug Design to Overcome the Problem. *Drug Discovery Today* **2012**, *17* (19−20), 1121−1126.

(23) Ayaz, P.; Andres, D.; Kwiatkowski, D. A.; Kolbe, C.-C.; Lienau, P.; Siemeister, G.; Lücking, U.; Stegmann, C. M. Conformational Adaption May Explain the Slow Dissociation Kinetics of Roniciclib (BAY 1000394), a Type I CDK Inhibitor with Kinetic Selectivity for CDK2 and CDK9. *ACS Chem. Biol.* **2016**, *11* (6), 1710−1719.

(24) Brynda, J.; Rezacova, P.; Fabry, M.; Horejsi, M.; Stouracova, R.; Sedlacek, J.; Soucek, M.; Hradilek, M.; Lepsik, M.; Konvalinka, J. A Phenylnorstatine Inhibitor Binding to HIV-1 Protease: Geometry, Protonation, and Subsite-Pocket Interactions Analyzed at Atomic Resolution. *J. Med. Chem.* **2004**, *47* (8), 2030−2036.

(25) Blizzard, T. A.; Dininno, F.; Morgan, J. D.; Chen, H. Y.; Wu, J. Y.; Kim, S.; Chan, W.; Birzin, E. T.; Yang, Y. T.; Pai, L.-Y.; Fitzgerald, P. M. D.; Sharma, N.; Li, Y.; Zhang, Z.; Hayes, E. C.; Dasilva, C. A.; Tang, W.; Rohrer, S. P.; Schaeffer, J. M.; Hammond, M. L. Estrogen Receptor Ligands. Part 9: Dihydrobenzoxathiin SERAMs with Alkyl Substituted Pyrrolidine Side Chains and Linkers. *Bioorg. Med. Chem. Lett.* **2005**, *15* (1), 107−113.

(26) Chien, E. Y. T.; Liu, W.; Zhao, Q.; Katritch, V.; Han, G. W.; Hanson, M. A.; Shi, L.; Newman, A. H.; Javitch, J. A.; Cherezov, V.; Stevens, R. C. Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist. *Science* **2010**, *330* (6007), 1091−1095.

(27) Segala, E.; Guo, D.; Cheng, R. K. Y.; Bortolato, A.; Deflorian, F.; Doré, A. S.; Errey, J. C.; Heitman, L. H.; IJzerman, A. P.; Marshall, F. H.; Cooke, R. M. Controlling the Dissociation of Ligands from the Adenosine A2A Receptor through Modulation of Salt Bridge Strength. *J. Med. Chem.* **2016**, *59* (13), 6470−6479.

(28) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55* (14), 6582−6594.

(29) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235−242.

(30) Chemical Computing Group Inc. *Molecular Operating Environment (MOE)*; 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, 2015.

(31) Katritch, V.; Fenalti, G.; Abola, E. E.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Allosteric Sodium in Class A GPCR Signaling. *Trends Biochem. Sci.* **2014**, *39* (5), 233−244.

(32) Fenalti, G.; Giguere, P. M.; Katritch, V.; Huang, X.-P.; Thompson, A. A.; Cherezov, V.; Roth, B. L.; Stevens, R. C. Molecular Control of δ-Opioid Receptor Signalling. *Nature* **2014**, *506* (7487), 191−196.

(33) LLC, S. *Schrödinger Release 2017-3: Maestro 11.3*; New York, 2017.

(34) Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L. OPM Database and PPM Web Server: Resources for Positioning of Proteins in Membranes. *Nucleic Acids Res.* **2012**, *40* (D1), D370−D376.

(35) LLC, S. *Schrödinger Release 2017-3: Desmond 5.1*; New York, 2017.

(36) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.

(37) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33−38.

(38) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13* (7), e1005659.

(39) Li, Y.; Zhang, J.; Gao, W.; Zhang, L.; Pan, Y.; Zhang, S.; Wang, Y. Insights on Structural Characteristics and Ligand Binding Mechanisms of CDK2. *Int. J. Mol. Sci.* **2015**, *16* (5), 9314−9340.

(40) Michino, M.; Beuming, T.; Donthamsetti, P.; Newman, A. H.; Javitch, J. A.; Shi, L. What Can Crystal Structures of Aminergic Receptors Tell Us about Designing Subtype-Selective Ligands? *Pharmacol. Rev.* **2015**, *67* (1), 198−213.

(41) Ghosh, A. K.; Osswald, H. L.; Prato, G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *J. Med. Chem.* **2016**, *59* (11), 5172−5208.

(42) Carpenter, B.; Lebon, G. Human Adenosine A2A Receptor: Molecular Mechanism of Ligand Binding and Activation. *Front. Pharmacol.* **2017**, *8* (Dec), 898.

(43) Knapp, B.; Ospina, L.; Deane, C. M. Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J. Chem. Theory Comput.* **2018**, *14* (12), 6127−6138.

(44) Caro, J. A.; Harpole, K. W.; Kasinath, V.; Lim, J.; Granja, J.; Valentine, K. G.; Sharp, K. A.; Wand, A. J. Entropy in Molecular Recognition by Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (25), 6563−6568.

(45) Tran-Nguyen, V.-K.; Da Silva, F.; Bret, G.; Rognan, D. All in One: Cavity Detection, Druggability Estimate, Cavity-Based Pharmacophore Perception and Virtual Screening. *J. Chem. Inf. Model.* **2019**, *59* (1), 573−585.

(46) Mortier, J.; Dhakal, P.; Volkamer, A. Truly Target-Focused Pharmacophore Modeling: A Novel Tool for Mapping Intermolecular Surfaces. *Molecules* **2018**, *23* (8), 1959.

Supporting Information

PyRod – Tracing Water Molecules in Molecular Dynamics Simulations

David Schaller†, Szymon Pach† and Gerhard Wolber†,*

† Pharmaceutical and Medicinal Chemistry, Freie Universität Berlin, Königin-Luise-Str. 2+4, 14195 Berlin, Germany

*Corresponding Author, e-mail: gerhard.wolber@fu-berlin.de

Table of contents
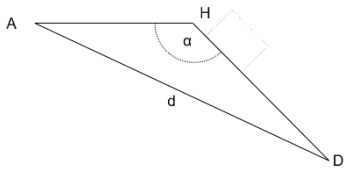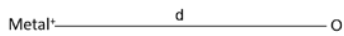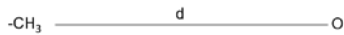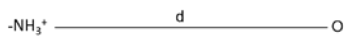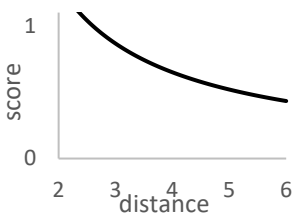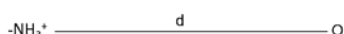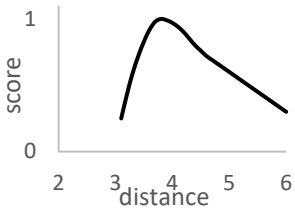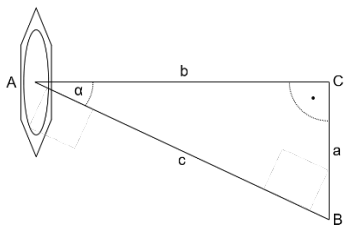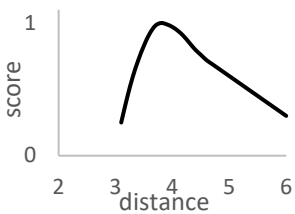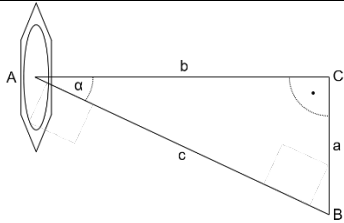
1. Feature Definition

Table S1: Chemical feature definition and functional protein groups.

protein chemical feature	selection criteria	corresponding ligand chemical features for interaction
hydrogen bond acceptor	• nitrogen, oxygen or sulfur with acceptor capabilities • imidazole nitrogen checked for protonation state	• hydrogen bond donor
hydrogen bond donor	• nitrogen, oxygen or sulfur with donor capabilities • checked for protonation state	• hydrogen bond acceptor
hydrophobic interaction	• uncharged carbon or sulfur atoms not bonded to oxygen or nitrogen	• hydrophobic interaction
positive ionizable	• N-terminus, lysine amine, arginine guanidine, histidine imidazole • checked for protonation state	• negative ionizable • aromatic interaction
negative ionizable	• C-terminus, aspartate carboxylate, glutamate carboxylate, cysteine thiolate, serine hydroxylate, threonine hydroxylate, tyrosine hydroxylate • checked for protonation state	• positive ionizable
aromatic interaction	• phenylalanine benzene, tyrosine benzene, tryptophan indole (benzene and pyrrole), histidine imidazole (checked for protonation)	• aromatic interaction • positive ionizable
metal	• name specified via config-file	• hydrogen bond acceptor

2. Scoring Functions

Table S2: Scoring functions for chemical features in PyRod. HB – hydrogen bond, HI – hydrophobic interaction, II – ionizable interaction, AI – aromatic interaction. * scored per grid point.

type	geometry	cutoffs	scoring function
HB		$d_{oxygen} \leq 3.2$ $d_{nitrogen} \leq 3.3$ $d_{sulfur} \leq 3.9$ $\alpha \leq 130°$	occupancy
Metal		$d \leq 3$	occupancy, part of hydrogen bond acceptor and negative ionizable
HI		$d \leq 5$	number of hydrophobic atoms scaled by buriedness
II		$d \leq 6$	occupancy scaled by distance
Cation-π*		$3.1 \leq d \leq 6$	occupancy scaled by distance
π-Cation*		$\alpha < 30°$ $3.1 < b \leq 6.0$	occupancy scaled by distance
$AI_{\pi\text{-stacked}}$*		$\alpha < 45°$ $3.3 \leq b \leq 4.7$ $a \leq 2.0$	occupancy scaled by distance
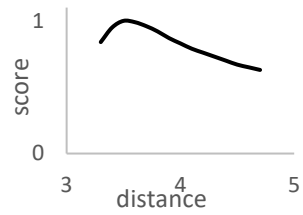
AI$_{\text{t-stacked}}$*
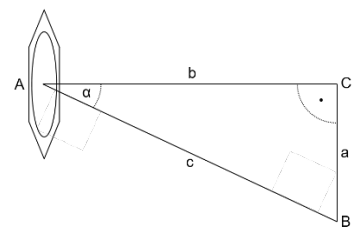
$\alpha < 45°$
$4.7 < b \leq 6.0$
$a \leq 0.5$
or
$\alpha \geq 45°$
$4.6 \leq a \leq 6.0$
$b \leq 0.5$

occupancy scaled by distance
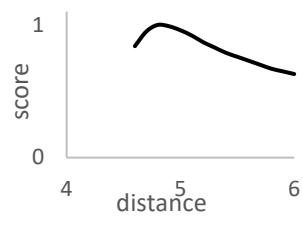
3. PyRod combinatorial library parameters

Table S3: Parameters used for combinatorial library generation.

		CDK2	D3R	HIV1P
minimal	independent chemical features	3	3	3
	hydrogen bonding features	1	1	2
	ionizable features	0	1	0
	aromatic features	0	0	0
	hydrophobic features	1	1	2
maximal	independent chemical features	5	5	5
	hydrogen bond features	4	4	4
	ionizable features	1	1	1
	aromatic features	2	2	2
	hydrophobic features	3	3	3
number of features in super pharmacophore		15	16	15
number of generated combinations		816	2441	588

4. Early enrichment factors

Table S4: Early enrichment factors for best performing pharmacophore model generated by PyRod and for docking results from DOCK 3.6[1] as benchmarked in the DUD-E publication[2].

	PyRod	Dock 3.6
CDK2	30.3	14
D3R	7.7	4
HIV1P	54.6	5
ERα	-	15
$A_{2A}R$	-	22

5. Equilibration
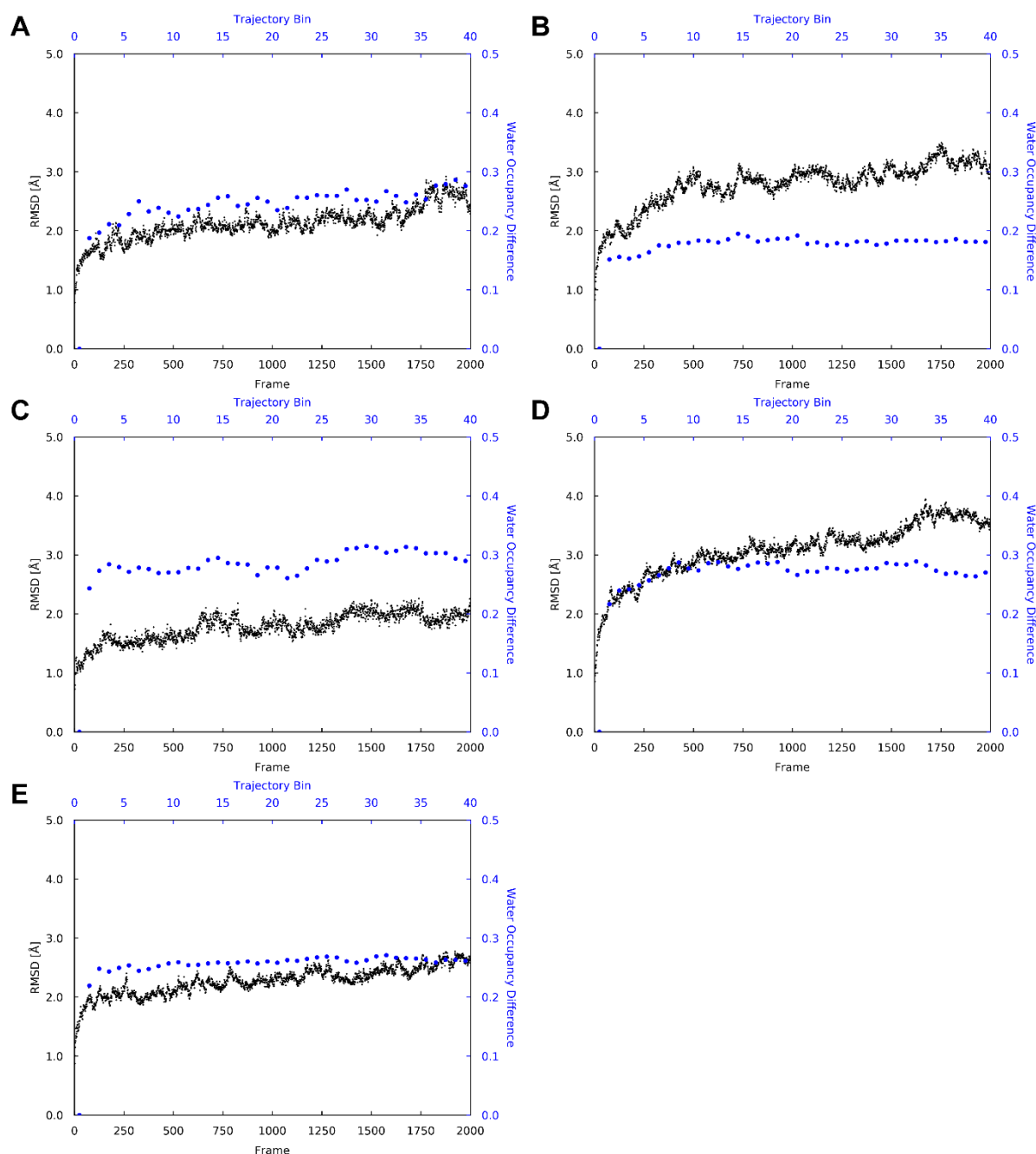
Figure S1: RMSD and water occupancy difference plots for one replication of each system tested, i.e. cyclin-dependent kinase 2 (5if1[3], **A**), dopamine D3 receptor (3pbl[4], **B**), HIV-1 protease (1nh0[5], **C**), estrogen receptor α (1xpc[6], **D**) and adenosine A_{2A} receptor (5iu4[7], **E**). RMSD of protein heavy atoms is shown in black and was calculated using VMD 1.9.2[37]. The water occupancy difference is shown in blue and was calculated as follows. The trajectory of each system was separated into trajectory bins of 50 frames. The first bin contains frames 1-50 and the last bin contains frames 1951-2000. PyRod was used for each trajectory bin to calculate water occupancies. The water occupancies of each bin were then compared grid-point wise to the water occupancies of the first bin, whereas a grid point pair was classified as different if the water occupancy values differ by more than 5. A water occupancy difference of 0.3 corresponds to a situation where 30 % of the grid points of that bin show a different water occupancy compared to the first bin.

6. References

(1) Coleman, R. G.; Carchia, M.; Sterling, T.; Irwin, J. J.; Shoichet, B. K. Ligand Pose and Orientational Sampling in Molecular Docking. *PLoS One* **2013**, *8* (10), e75992.

(2) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55* (14), 6582–6594.

(3) Ayaz, P.; Andres, D.; Kwiatkowski, D. A.; Kolbe, C.-C.; Lienau, P.; Siemeister, G.; Lücking, U.; Stegmann, C. M. Conformational Adaption May Explain the Slow Dissociation Kinetics of Roniciclib (BAY 1000394), a Type I CDK Inhibitor with Kinetic Selectivity for CDK2 and CDK9. *ACS Chem. Biol.* **2016**, *11* (6), 1710–1719.

(4) Chien, E. Y. T.; Liu, W.; Zhao, Q.; Katritch, V.; Han, G. W.; Hanson, M. A.; Shi, L.; Newman, A. H.; Javitch, J. A.; Cherezov, V.; Stevens, R. C. Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist. *Science* **2010**, *330* (6007), 1091–1095.

(5) Brynda, J.; Rezacova, P.; Fabry, M.; Horejsi, M.; Stouracova, R.; Sedlacek, J.; Soucek, M.; Hradilek, M.; Lepsik, M.; Konvalinka, J. A Phenylnorstatine Inhibitor Binding to HIV-1 Protease: Geometry, Protonation, and Subsite-Pocket Interactions Analyzed at Atomic Resolution. *J. Med. Chem.* **2004**, *47* (8), 2030–2036.

(6) Blizzard, T. A.; Dininno, F.; Morgan, J. D.; Chen, H. Y.; Wu, J. Y.; Kim, S.; Chan, W.; Birzin, E. T.; Yang, Y. T.; Pai, L.-Y.; Fitzgerald, P. M. D.; Sharma, N.; Li, Y.; Zhang, Z.; Hayes, E. C.; Dasilva, C. A.; Tang, W.; Rohrer, S. P.; Schaeffer, J. M.; Hammond, M. L. Estrogen Receptor Ligands. Part 9: Dihydrobenzoxathiin SERAMs with Alkyl Substituted Pyrrolidine Side Chains and Linkers. *Bioorg. Med. Chem. Lett.* **2005**, *15* (1), 107–113.

(7) Segala, E.; Guo, D.; Cheng, R. K. Y.; Bortolato, A.; Deflorian, F.; Doré, A. S.; Errey, J. C.; Heitman, L. H.; IJzerman, A. P.; Marshall, F. H.; Cooke, R. M. Controlling the Dissociation of Ligands from the Adenosine A2A Receptor through Modulation of Salt Bridge Strength. *J. Med. Chem.* **2016**, *59* (13), 6470–6479.

4.4   Exploiting Water Dynamics Enables Structure-Based Pharmacophore Searches Against MCHR1

After successful evaluation of our novel method (see section 4.3) we applied PyRod to MD simulations of an atomistic model of MCHR1. In the following study 3D pharmacophores were derived for MCHR1 and validated with a set of active and decoy molecules.

Contribution:

Conceptual design (90 %)
Computational experiments (100 %)
Visualization (100 %)
Manuscript preparation (90 %)

The following manuscript was submitted to Molecular Informatics.

molecular
informatics

PyRod Enables Rational Homology Model-Based Virtual Screening Against MCHR1

David Schaller[a] and Gerhard Wolber*[a]

Abstract: Several encouraging pre-clinal results highlight the melanin-concentrating hormone receptor 1 (MCHR1) as promising target for anti-obesity drug development. Currently however, experimentally resolved structures of MCHR1 are not available, which complicates rational drug design campaigns. In this study, we aimed at developing accurate, homology-model based 3D pharmacophores against MCHR1. We show that traditional approaches involving docking of small molecules are hindered by the flexibility of central binding pocket residues. Instead, we derived pharmacophore features from molecular dynamics simulations by employing our novel open-source software PyRod. The generated 3D pharmacophores were highly predictive returning up to 35 % of active molecules and showing an early enrichment (EF1) of up to 27.6. Furthermore, PyRod pharmacophores demonstrate higher sensitivity than ligand-based pharmacophores and deliver structural insights, which are key to rational lead optimization.

Keywords: MCHR1, PyRod, 3D pharmacophore, homology modeling, MD simulation

Obesity and overweight have progressed into major threats for human health causing 4 Mio deaths in 2015.[1] Beside bariatric surgery that is associated with several complications[2], pharmaceutical intervention in combination with lifestyle intervention proved to be the most promising treatment option for obesity.[3–5] However, currently approved anti-obesity agents lack efficacy and show severe or unpleasant side effects.[6]

The melanin-concentrating hormone receptor 1 (MCHR1) is a well characterized target for potential obesity treatment. Several rodent models of obesity showed encouraging results in knock-out experiments or in administration of MCHR1 antagonists. Unfortunately, these promising results could not be translated to human obesity treatment.[7] However, there is evidence that the simultaneous antagonism of MCHR1 and histamine H_3 receptor (H_3R) might result in a synergistic effect that could be beneficial in obesity treatment.[8] Also, we recently found three ligands that bind both receptors in the nanomolar activity range validating this target pair for rational multi-target drug design campaigns.[9]

Structure-based virtual screening campaigns employing atomistic models of the macromolecular target can be advantageous over ligand-based campaigns, since hits confirmed by in-vitro assays come with a potential binding hypothesis that can be exploited in subsequent lead optimization campaigns.[10] Especially multi-target drug design campaigns benefit from structural data, since lead molecules need to be optimized against multiple targets. Although the number of entries in the Protein Data Bank (PDB) is constantly increasing, many potential drug targets as well as validated drug targets still lack an experimentally resolved atomistic model.[11] In such situation, researchers often employ homology modeling, a method that is generating an atomistic model of the target of interest based on a closely related macromolecule.[12] However, performing structure-based virtual screening using homology models increases the chance for modeling artifacts, since even small modeling errors, such as a wrong side chain conformation essential for ligand binding, can impair docking performance.[13,14]

Molecular dynamics (MD) simulation can be used to address such artifacts and additionally, provide valuable information about the flexibility and thermodynamic properties of the system. [15–17] PyRod, a free and open-source Python software, combines the strength of MD simulations with structure-based 3D pharmacophore searches by analyzing the protein environment of water molecules in protein binding pockets and subsequently generates pharmacophore features for virtual screening.[18]

In this study, we aimed at generating highly predictive structure-based 3D pharmacophore models for virtual screening against MCHR1. A highly flexible hydrogen bond network involving three glutamine residues in the binding pocket of MCHR1 hindered the use of conventional workflows employing docking algorithms for pose prediction. Hence, we applied the free and open-source software PyRod, that analyzes the protein environment of water molecules in protein binding pockets throughout an MD simulation for pharmacophore feature placement. The presented workflow (Fig 1) yielded 3D pharmacophores that were highly successful in discriminating actives from decoys in a retrospective virtual screening campaign. Furthermore, they provide structural insights for binding

[a] *Pharmaceutical and Medicinal Chemistry, Freie Universität Berlin*
Königin-Luise-Strasse 2+4, 14195 Berlin, Germany
*e-mail: gerhard.wolber@fu-berlin.de, phone: +49 30 83852686

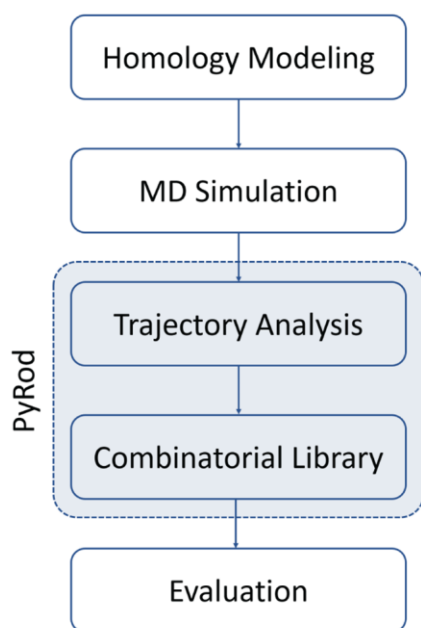to MCHR1 that are not obtainable by ligand-based pharmacophore modeling.

Figure 1. Workflow diagram for generating homology model-based 3D pharmacophores against MCHR1 with PyRod.

A sequence search found the crystal structure 4N6H[19] of the δ opioid receptor to be a suitable template for homology modeling of MCHR1 with a sequence similarity of 50.2 %. A homology model of MCHR1 in the inactive state was generated with MOE 2018[20]. The sodium ion complexed by $D^{2.50}$ was transferred from 4N6H into the MCHR1 homology model, since it was found to be structurally important for the inactive state of class A GPCRs.[21] A ramachandran plot analysis in MOE 2018 underlined the quality of the model with 95 % of dihedral angles located in the core region, 5 % in the allowed region and none outside.

The orthosteric pocket residues $Q^{3.36}$, $Q^{5.42}$ and $Q^{6.55}$ are highly flexible allowing various conformations and interact with several neighboring residues in a complex hydrogen bonding network (Fig 2). Such situations complicate docking studies, since their performance can already be affected by small changes in side chain conformations.[13,14] Thus, this homology model was subjected to molecular dynamics simulations with Desmond 5.1[22] to explore side chain conformations and to relax artifacts introduced through homology modeling. The trajectories were analyzed using PyRod 0.7.2[18] to identify potential hotspots for ligand binding and to generate 3D pharmacophores for virtual screening.

The PyRod software describes pharmacophoric binding pocket characteristics in form of dynamic molecular interaction fields (dMIFs) for common pharmacophore features including hydrogen bonds, ionizable and aromatic interactions as well as hydrophobic contacts. PyRod suggests favorable regions for hydrogen bonding and charged interactions close to $D^{3.32}$ and the sodium ion complexed by $D^{2.50}$ (Fig 3A). Additional hotspots for hydrogen bonding are located next to $Q^{3.36}$, $Q^{5.42}$ and $Q^{6.55}$ supporting our hypothesis

on the potential participation of these residues for ligand binding. Several hydrophobic residues are present in the orthosteric binding pocket favoring hydrophobic contacts above $D^{3.32}$, next to the sodium ion and close to the glutamines 3.36, 5.42 and 6.55 (Fig 3B). Sites for possible aromatic interactions can be found between residues $W^{6.48}$ and $F^{2.53}$ next to the sodium ion and in the upper part of the binding pocket next to extracellular loop residues R353 and F256.
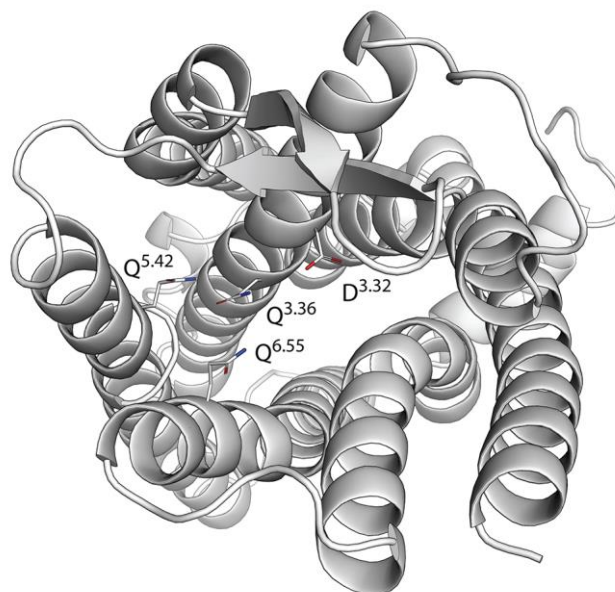
Figure 2. Top view into the binding pocket of the MCHR1 homology modeling. The three flexible glutamines can adapt various conformations.

PyRod generates a single super-pharmacophore by analyzing dMIFs for each pharmacophore feature type. However, this 3D pharmacophore consists of too many features for efficient virtual screening. Thus, prioritization and selection of pharmacophore features is mandatory for further processing and was performed by analyzing dMIFs for each respective pharmacophore feature type manually. Hotspots for interactions close to the sodium ions were ignored, since, to our knowledge, no ligands of GPCRs were reported to replace or interact with this sodium ion. Pharmacophore features were selected to have a high score according to the respective dMIF and to cover the orthosteric binding pocket. Additionally, pharmacophore features were included that are located close to extracellular loops, since these regions were found to frequently contribute to ligand binding.[23] The focused 3D pharmacophore model consists of 15 features (Fig 4A), i.e. two positive ionizable interaction features and two associated hydrogen bond donors pointing towards $D^{3.32}$, four hydrogen bonding features close to glutamines 3.36, 5.42 and 6.55, two hydrogen bond acceptors close to the extracellular loops of MCHR1, four hydrophobic features covering both pockets next to the three glutamines and above $D^{3.32}$, and one aromatic feature next to R353.
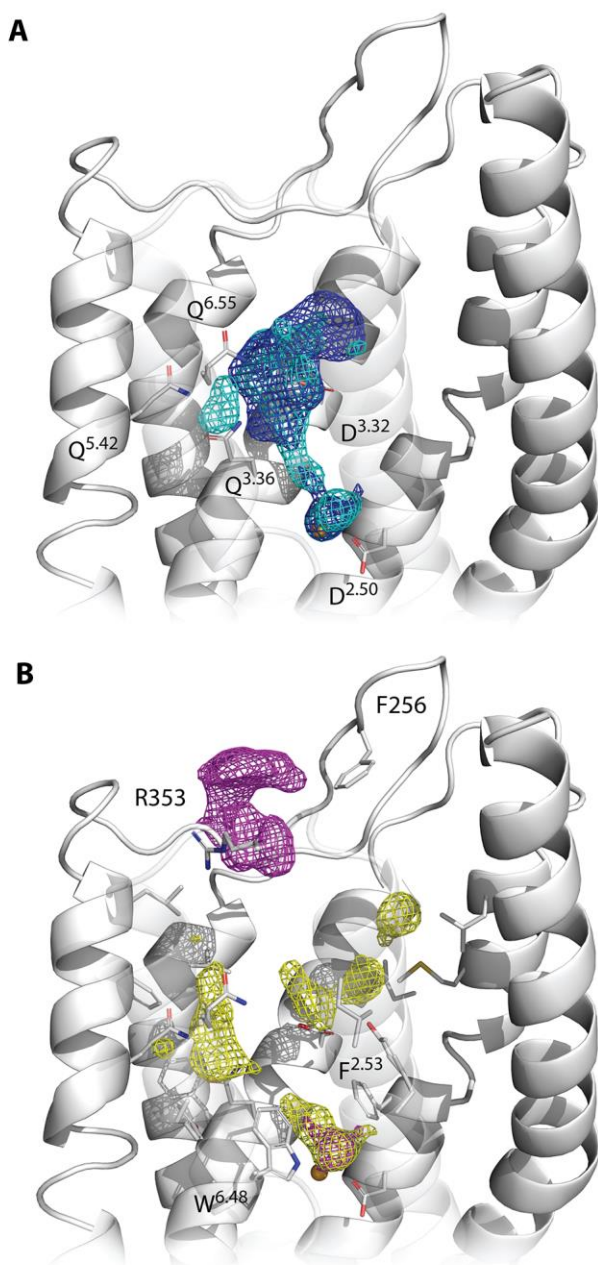
Figure 3. PyRod analysis of MCHR1 orthosteric binding pocket. Depicted dMIFs represent favorable regions for (A) hydrogen bonding (cyan, cutoff=27.8) and positive ionizable interactions (blue, cutoff=25.3) as well as for (B) hydrophobic (yellow, cutoff=111.8) and aromatic features (magenta, cutoff=15.3). Cutoffs were chosen based on the half maximum of the respective feature class. Transmembrane helices 6 and 7 were set transparent to allow better visualization.

The focused 3D pharmacophore was subjected to combinatorial processing with PyRod 0.7.2.[18] Feature combinations were restricted to 3D pharmacophores with minimal 3 and maximal 5 independent features to reduce the combinatorial space. Additionally, each 3D pharmacophore must contain one positive ionizable feature to further limit combinatorial space and to focus on ligands carrying a positive charge which would be beneficial for potential binding to H_3R. This procedure resulted in 1136 different 3D pharmacophores against MCHR1. 3D pharmacophores were evaluated with

LigandScout 4.2[24] for discrimination of a diverse set of 100 actives retrieved from the ChEMBL 24 database[25] and 6350 matched decoys from the DUD-E server[26].

Altogether, 62 3D pharmacophores found at least 5 % of the MCHR1 active set, which was the criteria to advance to the computational expensive decoy screening (Fig 4B). The results from actives and decoys screening were used for calculation of the early enrichment factor ($EF_{1\%}$). The 3D pharmacophore with the highest true positive hit rate finds 35 % of screened actives and consists of one positive ionizable feature next to $D^{3.32}$, one hydrophobic feature above three glutamines 3.36, 5.42 and 6.55, as well as one hydrogen bond acceptor feature close to the extracellular loops (Fig 4C, supporting information Fig S1A). However, this 3D pharmacophore achieves only considerably weak early enrichment ($EF_{1\%}$ = 4.0). In contrast, the 3D pharmacophore with the best enrichment ($EF_{1\%}$ = 27.6) also carries an aromatic feature next to R353 and additionally has the hydrophobic feature located above $D^{3.32}$, but only picks 6 % of the active molecules (Fig 4D, supporting information Fig S1B).

A popular approach in the absence of experimentally resolved atomistic models is the generation of 3D pharmacophores from alignments of known active molecules. Hence, we were interested if PyRod pharmacophores can achieve a similar performance compared to ligand-based pharmacophores. The complete active set, containing 695 unique MCHR1 ligands ,was clustered with LigandScout 4.2[24] resulting in 19 clusters comprising of at least 10 molecules. Each of these clusters was employed to generate a shared-feature pharmacophore. In total, 12 pharmacophores contained the important positive ionizable feature and were evaluated for early enrichment and retrieval of known actives (Fig 4B, supporting information Fig S1C). All tested ligand-based shared-feature pharmacophores show very high early enrichment factors of up to 64.5. However, the ligand-based pharmacophores are not sensitive returning at most 6 % of actives from the test set. Furthermore, such 3D pharmacophores lack any information about the interactions with the receptor, which is essential to rational lead optimization.

This is the first study applying PyRod on MD simulations of a homology model. By employing PyRod, we were able to generate several 3D pharmacophores against MCHR1 that are highly successful in discriminating active MCHR1 ligands from decoys. The 3D pharmacophore generation is thereby not dependent on error-prone docking studies in homology models but instead exploits water dynamics from MD simulations. Hits identified with these structure-based 3D pharmacophores hold the information of a binding hypothesis that can be used for subsequent rational lead optimization. Furthermore, we show that PyRod pharmacophores present an attractive alternative to ligand-based pharmacophores that heavily dependent on correct ligand conformations as well as their alignment, and additionally, lack information essential for further lead optimization.[27] These characteristics render PyRod pharmacophores highly valuable tools for hit identification and optimization in anti-obesity drug design campaigns against MCHR1. Also, the presented workflow (Fig 1) can be easily transferred to other

3

projects that aim at performing homology model-based

virtual screening campaigns.
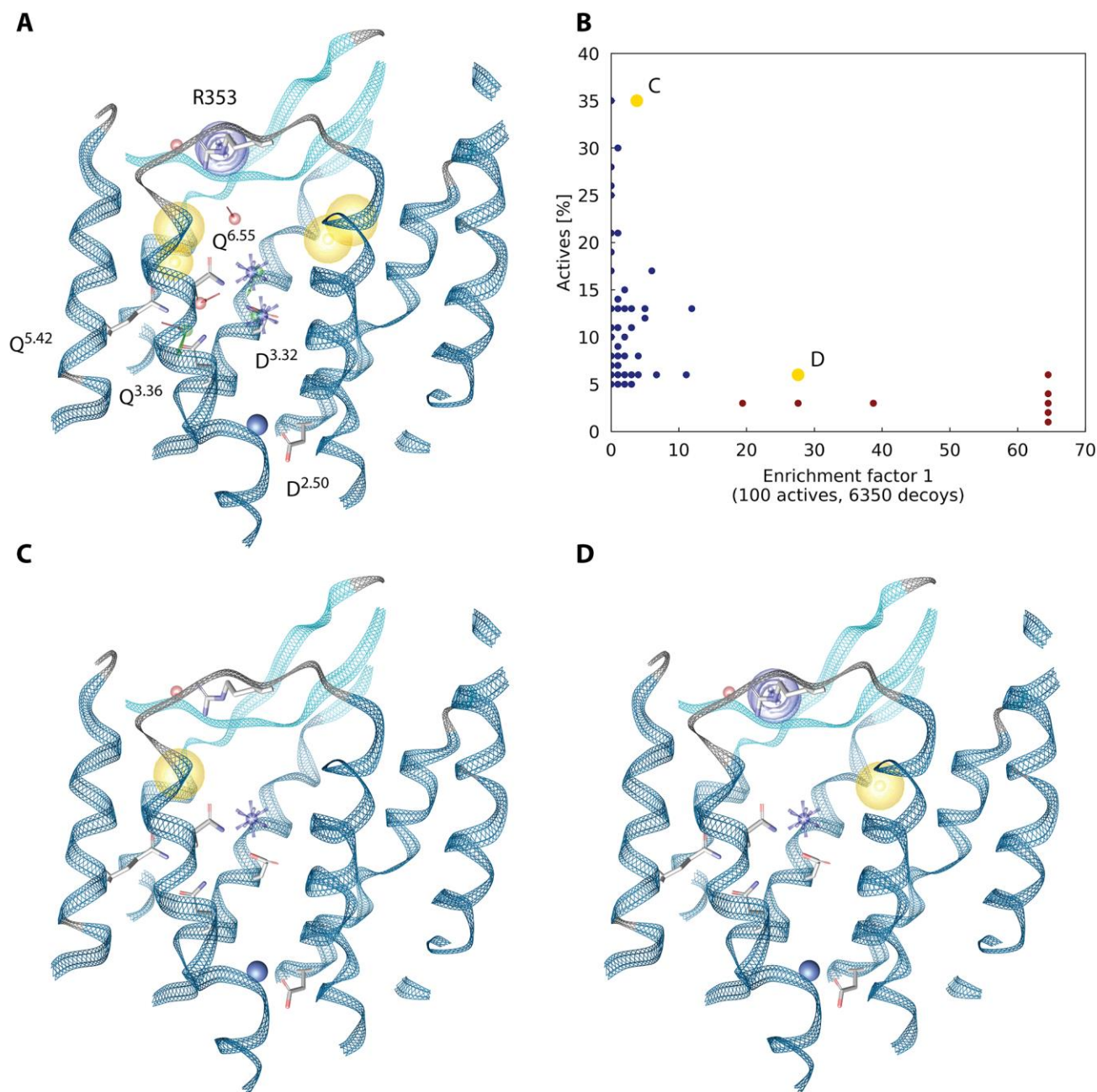
Figure 4. 3D pharmacophores and their performance. (A) Focused 3D pharmacophore from PyRod used for combinatorial processing. (B) Evaluation of 3D pharmacophores against a MCHR1 test set. The blue and yellow dots represent the performance of PyRod pharmacophores, red dots represent the performance of ligand-based shared feature pharmacophores generated with LigandScout 4.2. (C, D) 3D pharmacophores identifying the most actives from MCHR1 test set and showing the highest early enrichment respectively. Exclusion volumes were not depicted for the sake of clarity. Blue star – positive ionizable, yellow sphere – hydrophobic contact, purple ring – aromatic interaction, red arrow – hydrogen bond acceptor.

Experimental Section

A template search using the GPCRdb[28] and subsequent analysis in MOE 2018[20] revealed the high resolution crystallographic structure 4N6H[19] of the δ opioid receptor as suitable template for generating a homology model of MCHR1. The amino acid sequence of human MCHR1

(Q99705) was retrieved from Uniprot[29] and aligned to the crystallographic structure 4N6H in MOE 2018 according to the proposed alignment from GPCRdb (supporting information Fig S2). The aligned sequences show a sequence identity of 29.4 % and a sequence similarity of 50.2 %.

Employing this alignment, a homology model of human MCHR1 was generated based on 10 main chain models with 10 side chain samples per main chain

model at 300 K in MOE 2018. Automatic model refinement was disabled. The structurally important sodium ion and 5 water molecules were transferred from the template structure 4N6H. The side chain conformation of S195 was refined to allow correct complexation of the sodium ion (supporting information Fig S3). Atom clashes were sequentially minimized with OPLS-AA force field[30] implemented in MOE 2018. Protonation was performed using the Protonate3D tool in MOE 2018.

The homology model of MCHR1 was subjected to molecular dynamics simulation. Chain breaks were capped with NME and ACE in MOE 2018[20]. The receptor was oriented using the PPM server[31] for subsequent membrane placement in a POPC bilayer using Maestro 11.3[32] and solvation in a orthorhombic box of TIP4P water with 10 Å padding containing 0.15 M NaCl. In total, 10 replica of 30 ns MD simulations were performed using Desmond 5.1[22]. Frames were saved every 10 ps resulting in 3000 frames per simulation. The pbc wrap functionality implemented in VMD 1.9.3[33] was employed to center the receptor in the periodic boundary box and the RMSD Trajectory Tool to align the trajectory on the heavy atoms of the protein backbone of the first frame.

The test grid component of PyRod 0.7.2[18] was used to identify appropriate parameters for grid placement. The identified parameters result in cubic grids with an edge length of 30 Å spanning the orthosteric binding pocket of MCHR1 (supporting information Fig S4). The last 10 ns of each simulation were analyzed using the trajectory pharmacophore combo of PyRod 0.7.2 with default parameters resulting in the generation of dynamic molecular interaction fields describing pharmacophoric binding pocket characteristics as well as a super pharmacophore describing potential interaction sites with the receptor.

The CHEMBL 24 database[25] was used to retrieve activity data for MCHR1 (CHEMBL344). Ligands were filtered for molecular weight (≤ 700), confidence score (≥ 9), standard relation (=), standard value (≤ 10) standard units (nM) and standard type (K_i, K_d, IC_{50} or EC_{50}). RDKit[34] nodes implemented in KNIME 3.7.1[35] were used to remove molecules with unspecified stereo centers and to remove duplicates, whereat binding data was preferred over functional data and more recent data points were preferred over older. This procedure resulted in 695 unique ligands of MCHR1.

MOE 2018[20] was used to identify the dominant protonation state at pH 7 and Corina 3.00[36] to generate a low-energy 3D conformation. The RDKit diversity picker was employed in KNIME 3.7.1 to pick 100 diverse active ligands. The DUD-E server[26] was used to generate decoys for the selected diverse ligands. In total, 6350 decoys were retrieved from DUD-E server, protonated at pH 7 in MOE 2018 and an initial conformation was generated with Corina 3.00. By employing iCon implemented in idbgen from LigandScout 4.2[24] 25 conformations were generated for each of the molecules in the active and decoy sets for later 3D pharmacophore evaluation.

LigandScout 4.2 was employed to visualize and analyze the previously generated dMIFs guiding the selection of pharmacophore features from the super pharmacophore for combinatorial library generation with PyRod 0.7.2[18]. Fifteen features were selected and combined to 1136 different 3D pharmacophores. The combinatorial space was limited by restricting 3D pharmacophores to contain 3 – 5 independent features, 1 – 3 hydrogen bonds, 0 – 1 aromatic interaction and exactly 1 ionizable interaction. Each 3D pharmacophore was evaluated with LigandScout 4.2 for discrimination of actives from decoys which were generated as already described.

Ligand-based shared-feature pharmacophores were generated in LigandScout 4.2. All 695 unique MCHR1 ligands were clustered and clusters comprising of at least 10 molecules were subjected to shared-feature pharmacophore generation. 3D pharmacophores containing a positive ionizable feature were evaluated for early enrichment factor and retrieval of actives as already described.

References

[1]   S. Banerjee, A. Mitra, *N. Engl. J. Med.* **2017**, *377*, 13–27.

[2]   J. D. Birkmeyer, J. F. Finks, A. O'Reilly, M. Oerline, A. M. Carlin, A. R. Nunn, J. Dimick, M. Banerjee, N. J. O. Birkmeyer, *N. Engl. J. Med.* **2013**, *369*, 1434–1442.

[3]   N. Finer, S. Finer, R. P. Naoumova, *Am. J. Clin. Nutr.* **1992**, *56*, 195S-198S.

[4]   B. Richelsen, S. Tonstad, S. Rossner, S. Toubro, L. Niskanen, S. Madsbad, P. Mustajoki, A. Rissanen, *Diabetes Care* **2007**, *30*, 27–32.

[5]   T. A. Wadden, P. Hollander, S. Klein, K. Niswender, V. Woo, P. M. Hale, L. Aronne, *Int. J. Obes.* **2013**, *37*, 1443–1451.

[6]   G. Srivastava, C. M. Apovian, *Nat. Rev. Endocrinol.* **2018**, *14*, 12–24.

[7]   T. Högberg, T. M. Frimurer, P. K. Sasmal, *Bioorganic Med. Chem. Lett.* **2012**, *22*, 6039–6047.

[8]   G. S. Parks, N. D. Olivas, T. Ikrar, N. M. Sanathara, L. Wang, Z. Wang, O. Civelli, X. Xu, *J. Physiol.* **2014**, *592*, 2183–96.

[9]   D. Schaller, S. Hagenow, G. Alpert, A. Naß, R. Schulz, M. Bermudez, H. Stark, G. Wolber, *ACS Med. Chem. Lett.* **2017**, *8*, 648–653.

[10]   G. Sliwoski, S. Kothiwale, J. Meiler, E. W. Lowe, *Pharmacol. Rev.* **2014**, *66*, 334–395.

[11]   H. M. Berman, *Nucleic Acids Res.* **2000**, *28*, 235–242.

[12]   T. Schmidt, A. Bergner, T. Schwede, *Drug Discov. Today* **2014**, *19*, 890–897.

[13]   I. Kufareva, V. Katritch, R. C. Stevens, R. Abagyan, *Structure* **2014**, *22*, 1120–1139.

[14]   Z. Miao, Y. Cao, *Sci. Rep.* **2016**, *6*, 37024.

[15]   L. Heo, M. Feig, *Proc. Natl. Acad. Sci.* **2018**, *115*, 13276–13281.

[16]   J. Mortier, C. Rakers, M. Bermudez, M. S. Murgueitio, S. Riniker, G. Wolber, *Drug Discov. Today* **2015**, *20*, 686–702.

[17]   D. Schaller, D. Sribar, T. Noonan, L. Deng, T. N. Nguyen, S. Pach, D. Machalz, M. Bermudez, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, accepted.

[18]   D. Schaller, S. Pach, G. Wolber, *J. Chem. Inf. Model.* **2019**, *59*, 2818–2829.

[19]   G. Fenalti, P. M. Giguere, V. Katritch, X.-P. Huang, A. A. Thompson, V. Cherezov, B. L. Roth, R. C. Stevens, *Nature*

molecular
informatics

2014, *506*, 191–196.

[20]   Chemical Computing Group Inc., **2018**, Molecular Operating Environment (MOE).

[21]   V. Katritch, G. Fenalti, E. E. Abola, B. L. Roth, V. Cherezov, R. C. Stevens, *Trends Biochem. Sci.* **2014**, *39*, 233–244.

[22]   Schrödinger LLC, **2017**, Schrödinger Release 2017-3: Desmond 5.1.

[23]   M. Wheatley, D. Wootten, M. Conner, J. Simms, R. Kendrick, R. Logan, D. Poyner, J. Barwell, *Br. J. Pharmacol.* **2012**, *165*, 1688–1703.

[24]   G. Wolber, T. Langer, *J. Chem. Inf. Model.* **2005**, *45*, 160–169.

[25]   A. Gaulton, A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrián-Uhalte, et al., *Nucleic Acids Res.* **2017**, *45*, D945–D954.

[26]   M. M. Mysinger, M. Carchia, J. J. Irwin, B. K. Shoichet, *J. Med. Chem.* **2012**, *55*, 6582–6594.

[27]   J. Shim, A. D. MacKerell, Jr., *Medchemcomm* **2011**, *2*, 356.

[28]   V. Isberg, S. Mordalski, C. Munk, K. Rataj, K. Harpsøe, A. S. Hauser, B. Vroling, A. J. Bojarski, G. Vriend, D. E. Gloriam, *Nucleic Acids Res.* **2016**, *44*, D356–D364.

[29]   UniProt Consortium, *Nucleic Acids Res.* **2015**, *43*, D204–D212.

[30]   W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

[31]   M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, *Nucleic Acids Res.* **2012**, *40*, D370–D376.

[32]   S. LLC, **2017**.

[33]   W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33–38.

[34]   RDKit, "Open-source cheminformatics; http://www.rdkit.org.," **2018**.

[35]   M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, B. Wiswedel, in *Data Anal. Mach. Learn. Appl.*, Springer, Berlin, Heidelberg, **2008**, pp. 319–326.

[36]   J. Sadowski, J. Gasteiger, G. Klebe, *J. Chem. Inf. Model.* **1994**, *34*, 1000–1008.

6

Supporting Information

PyRod Enables Rational Homology Model-Based Virtual Screening Against MCHR1

David Schaller[a] and Gerhard Wolber*[a]

[a] Pharmaceutical and Medicinal Chemistry, Freie Universität Berlin, Königin-Luise-Straße 2+4, 14195 Berlin, Germany

*e-mail: gerhard.wolber@fu-berlin.de, phone: +49 30 83852686

Table of contents

1. ROC statistics

A
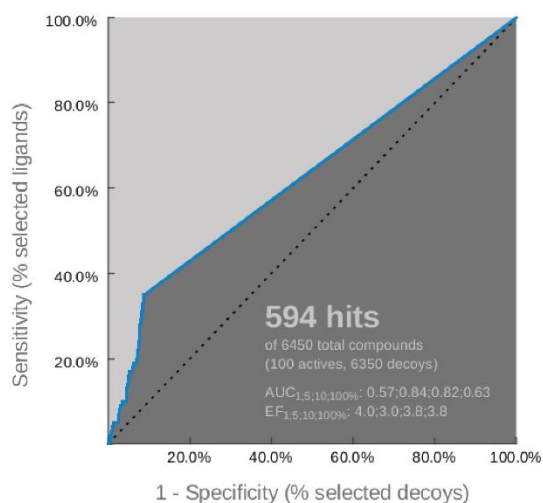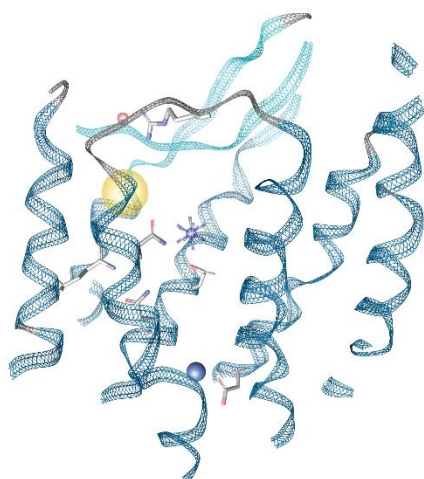
B
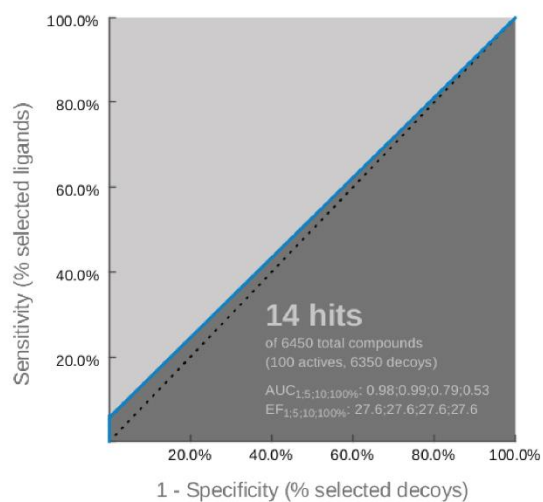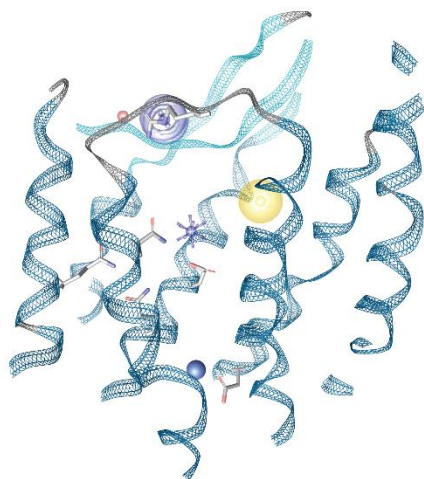
C
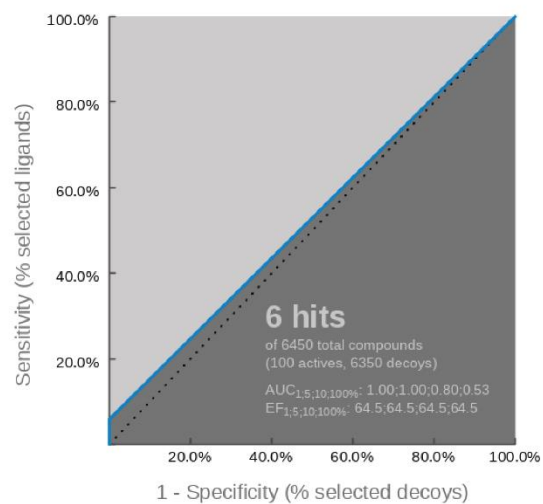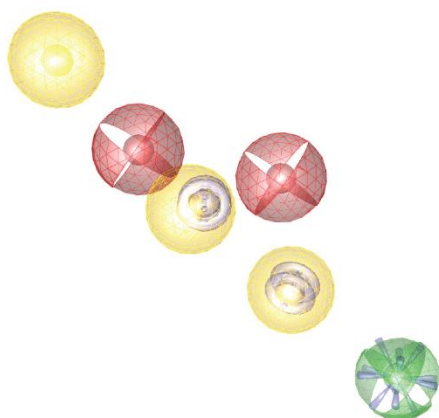
Figure S1: ROC statistics for selected 3D pharmacophores generated with PyRod (A, B) and with the ligand-based mode in LigandScout 4.2 (C). Exclusion volumes are not depicted for the sake of clarity. Blue star – positive ionizable, yellow sphere – hydrophobic contact, purple ring – aromatic interaction, red arrow and sphere – hydrogen bond acceptor, green sphere - hydrogen bond donor.

2. Sequence Alignment

```
              <---------------H1----------------->              <--------------
MCHR1         PRTGSISYINIIMPSVFGTICLLGIIGNSTVIFAVVKKSKLHWCNNVPDIFIINLSVVDL
DOR (4N6H)    ARSASSLALAIAITALYSAVCAVGLLGNVLVMFGIVRYTKM---KTATNIYIFNLALADA
              *..*     : * : :::.::* :*::** *:*.:*: :*:     .. .*:*:**: ..*
                                                                <-ICL1-->

              H2------------>         <--------------H3--------------->
MCHR1         LFLLGMPFMIHQLMGNGVWHFGETMCTLITAMDANSQFTSTYILTAMAIDRYLATVHPIS
DOR (4N6H)    LATSTLPFQSAKYLME-TWPFGELLCKAVLSIDYYNMFTSIFTLTMMSVDRYIAVCHPVK
              *     :**    : : : * *** :*. : ::*  . *** : ** *::***:*..**:.
                             <ECL1->                              <---

              <-----------H4---------->             <--------
MCHR1         STKFRKPSVATLVICLLWALSFISITPVWLYARLIPFPGGAVGCGIRLPNP-DTDLYWFT
DOR (4N6H)    ALDFRTPAKAKLINICIWVLASGVGVPIMVMAVTRPRD-GAVVCMLQFPSPSWYWDTVTK
              : .**.*: *.*:    :*.*:     .*: : *    *    *** * :::*.*
              ICL2->                          <-------ECL2------->

              --------H5----------------->        <-----------------H6----
MCHR1         LYQFFLAFALPFVVITAAYVRILQRMTSSVAPASQRSIRLRTKRVTRTAIAICLVFFVCW
DOR (4N6H)    ICVFLFAFVVPILIITVCYGLMLLRLRSVRLLSGSKEKDRSLRRITRMVLVVVGAFVVCW
              :  *::**.:*:::**..*  :* *: *    :..:.    :*:** .:.: .*.***
                                                <ICL3->

              ------------>      <------------H7------------> <----H8---->
MCHR1         APYYVLQLTQLS  ISRPTLTFVYLYNAAISLGYANSCLNPFVYIVLCETFRKRLVLSVK
DOR (4N6H)    APIHIFVIVWTLVDIDRRDPLVVAALHLCIALGYANSSLNPVLYAFLDENFKRCFRQLCR
              ** ::: :     * .*     .*   . .:*:******.***.:* .* *.*:: :     .
                                <ECL3->

MCHR1         PAAQ
DOR (4N6H)    KPCG
```

Figure S2: The depicted sequence alignment was used for homology modeling of human MCHR1. The template sequence is highlighted according to structural properties. Red sections represent helices, yellow sections represent β-sheets and underlined cysteines are involved in a disulfide bond. Furthermore, the sequence alignment contains information about the naming of helices and loops as well as the sequence similarity. (H1) - helix 1, (ICL1) - intracellular loop 1, (ECL1) - extracellular loop 2, (*) - identical residues, (:) - residues with high similarity, (.) - residues with low similarity.

3. Optimization of Sodium Coordination
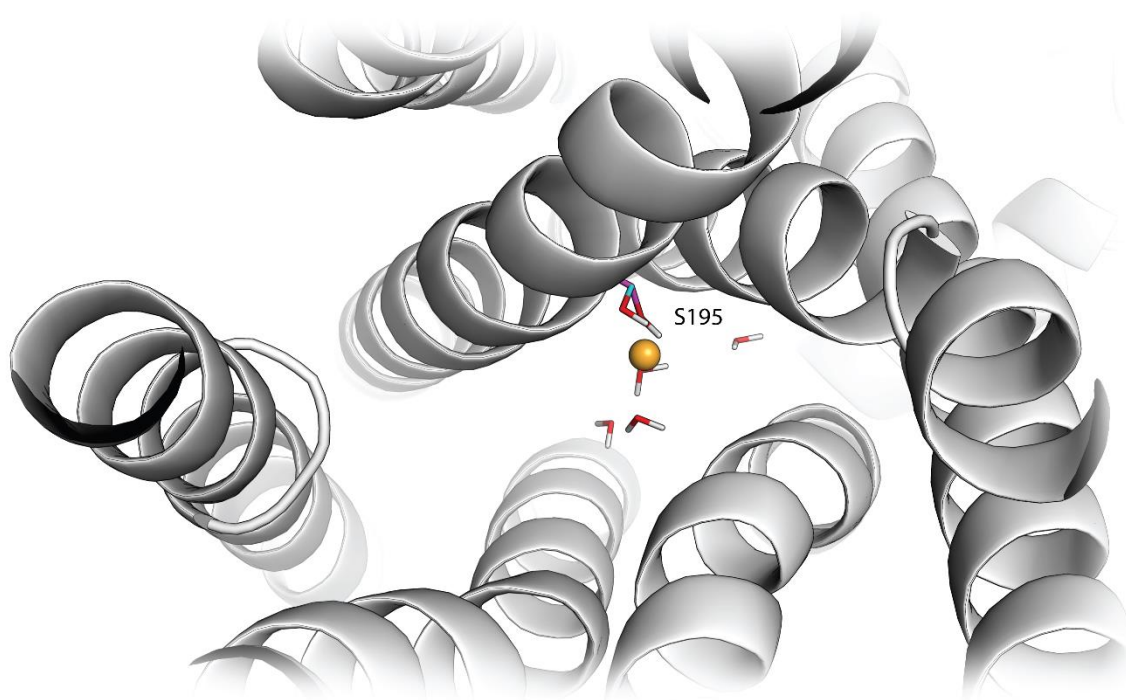
Figure S3: Refinement of the homology model of MCHR1. The conformation of S195 was altered to allow correct complexation of the important sodium ion highlighted in orange. Carbon atoms prior refinement are colored in cyan, carbon atoms after refinement in magenta.

Discussion

After decades of research, health systems still lack an appropriate strategy to stop the obesity epidemic. Rational drug design campaigns for compounds able to modulate multiple obesity-relevant targets present a promising approach to overcome the low efficacy and severe side effect of currently available medications. The increasing public availability of small molecule bioactivity data presents a unique opportunity to identify targets for multi-target drug development. However, the lack of experimentally resolved atomistic models limits our understanding of critical ligand-protein interactions and consequently hampers the complex lead optimization towards the desired activity profile against multiple obesity-relevant targets.

5.1 Identification of H$_3$R/MCHR1 as Promising Target Pair for Obesity Treatment

The reported data mining campaign (see section 4.1) assessed the similarity of publicly available small molecule activity data for 39 obesity-relevant drug targets revealing 20 target pairs that show binding of similar ligands towards different targets. Taylor-Butina [140] clustering was performed to prioritize target pairs by the number of similar ligand clusters. The maximum of three similar ligand clusters was identified for the target pairs H$_3$R/MCHR1, H$_3$R/μ$_1$ opioid receptor (μ$_1$OR) and MCHR1/serotonin receptor 2C (5HTR$_{2C}$). Notably, all these targets belong to the GPCR protein family.

Ligands of GPCRs can either activate (agonist) or inactivate (antagonist or inverse agonist) receptor signaling. This mechanism of action is essential to the pharmacological response. If activation of a certain receptor is attributed to anti-obesity effects, inactivation of this receptor will be ineffective or even promote obesity. Since these characteristics are not annotated in public databases, ligands were manually evaluated for their mechanism of action by reviewing the primary literature. This resulted in the identification of 11 target pairs with antiobese mechanism of action for both elements of the target pair. Five target pairs were found, in which one element had a conflictive mechanism of action and thus could be important as potential off-target. Unfortunately, the mechanism of action for ligands of five target pairs were not prop-

erly reported in the primary literature and thus could not be further evaluated.

Our data mining approach identified H_3R/MCHR1 as the most interesting target pair unveiling three similar ligand clusters with antiobese mechanism of action for both elements of the target pair. A literature research only yielded two studies reporting compounds with activity towards both receptors [89, 141]. However, the studied compounds show low potency for H_3R, since the authors were mainly interested into MCHR1 ligands.

A shape-based screening campaign with subsequent *in vitro* validation found three small molecules with prior unknown multi-target character against H_3R and MCHR1. The most active compound shows a balanced profile with two digit nanomolar activity at both receptors. The ligand efficiency of 0.34 characterizes this compound as good starting point for further optimization [142]. However, the lipophilicity-corrected ligand efficiency of 12.40 indicates potential promiscuity, since high lipophilicity is often associated with unspecific binding to many targets [142]. Indeed, closely related compounds were already found to show affinity towards $5HTR_{2C}$. Notably, $5HTR_{2C}$ was also identified by the presented data mining workflow to be a potential off-target for MCHR1 antagonists in anti-obesity drug development.

Both H_3R and MCHR1 were extensively studied for the development of potential anti-obesity drugs. However, drug candidates for both receptors failed to proof efficacy in clinical studies. A recent study might explain why the efforts in targeting H_3R for anti-obesity treatment did not lead to an effective treatment yet. Parks and colleagues found that H_3R antagonists increase the expression of orexigenic MCH which may abolish anti-obesity effects of H_3R antagonist [143]. Hence, a concurrent inhibition of MCHR1 might be essential for efficacy and lead to a synergistic effect.

Taken together, the applied data mining approach revealed several interesting target pairs binding similar ligands. The simultaneous antagonism of H_3R and MCHR1 is a promising approach towards an effective anti-obesity treatment. Three small molecules were identified that are able modulate both receptors simultaneously. The included off-target analysis as well as closely related compounds of the most active ligand indicate potential antagonistic activity at $5HTR_{2C}$ which could decrease the efficacy and should be considered in further development.

5.2 Advancing Rational Drug Design Against H_3R and MCHR1

Optimizing the activity of lead molecules against multiple targets is challenging, since modifications increasing the activity for one target are likely to decrease the activity for the other. Such lead optimization campaigns can dramatically benefit from including atomistic models of the targets, since the synthesis of analogues can be rationally

restricted to those showing favorable interactions to both targets. However, experimentally resolved atomistic models for H_3R and MCHR1 are currently not available. Thus, homology models have to be employed to advance our understanding of ligand binding which can ultimately promote lead optimization.

5.2.1 Ligand-Guided Homology Modeling of H_3R

Investigations utilizing homology models come with a substantial risk of modeling artifacts, since erroneous conformations of side chains critical for ligand interactions can dramatically impair docking results [144, 145]. To overcome this challenge, the presented workflow (see section 4.2) evaluates homology models of H_3R to allow a critical interaction between docked ligands and $D^{3.32}$ which is well characterized for aminergic receptors and can be observed in several experimentally resolved atomistic models [146, 147].

An analysis of docking poses for 1000 homology models of H_3R revealed that the critical charged interaction towards $D^{3.32}$ can be observed for 25 % of generated homology models in only less than 10 % of docking poses. This indicates structural problems that hinder a correct ligand placement. Noteworthy, 7 models were identified that do not allow a single docking pose with an interaction towards $D^{3.32}$. In contrast, the best ranked homology model allows the critical interaction in 83.5 % of docking poses. Thus, ligand data together with a single, yet reliable interaction could be utilized to eliminate homology models unlikely to perform well in further computational studies.

The difference of best and worst performing models could thereby not be explained by frequently employed scoring metrices, including dihedral angle outliers, heavy atom clashes or scores by homology model evaluation programs. A detailed comparison of side chain conformations of ten best and ten worst ranked models highlight the side chain conformation of $E^{5.46}$ to be critical for ligand placement. In the best performing homology models $E^{5.46}$ is pointing inside the binding pocket and is involved in a hydrogen bond with the docked ligands. In contrast, the worst performing models share a side chain conformation of $E^{5.46}$ that is pointing towards the lipophilic membrane. Also, no side chains are present in this region to compensate for the negative charge of $E^{5.46}$. These results support other studies suggesting an important role of $E^{5.46}$ for ligand binding [148–150]. The second most different side chain conformation was found for $L^{7.42}$. However, its different conformation could not be related to the docking results.

The best performing homology model was employed for a structure-based virtual screening campaign for novel ligands of H_3R starting with docking of 10 diverse known antagonists. Remarkably, the previously identified important residue $E^{5.46}$ is not always involved in interactions with the docked antagonists. However, one of the docked antagonists only contains a single hydrogen bond donor. Also, mutational studies for

closely related histamine H1 receptor indicate an importance of residues at position 5.46 only for some ligands [151, 152]. Thus, docking poses were selected which interact with critical residue $D^{3.32}$. A 3D pharmacophore-based screening campaign employing selected docking poses of 3 known antagonist with subsequent *in vitro* validation revealed two novel ligands of H_3R in the nanomolar activity range. These hits deliver rationalized binding modes that can ultimately aid the lead optimization process.

Taken together, the presented ligand-guided homology model workflow was able to detect side chain conformations important to ligand docking without the use of error prone docking scores. Instead, a single, yet reliable interaction was exploited to guide the selection of a homology model for virtual screening. The *in vitro* validation of virtual screening hits revealed two novel nanomolar inhibitors of H_3R with rationalized binding modes.

5.2.2 PyRod Enables Structure-Based Pharmacophore Searches Against MCHR1

To our knowledge no high quality mutations are available for MCHR1 which was essential to the ligand-guided homology workflow applied to H_3R. Also, three highly flexible glutamines $Q^{3.36}$, $Q^{5.42}$ and $Q^{6.55}$ hinder docking studies which can be affected by subtle conformational changes in the binding pocket. Hence, we shifted our focus towards the dynamic nature of protein binding sites which ultimately let to the development of PyRod, a program capable of generating structure-based pharmacophores from molecular dynamics simulation (see section 4.3).

The implemented routines in PyRod trace water molecules throughout a trajectory of protein-water conformations and analyze the protein environment of water molecules in the binding site. The gathered information reports interactions between water molecules and protein residues, e.g. hydrogen bonds, which could be exploited by small molecules to perform interactions with the protein. Furthermore, hydrophobic environments are captured which might be filled by complementing ligand moieties. Importantly, the scoring scheme highlights areas with conformationally restricted water molecules, whose replacement by a ligand is more likely to increase the entropy of the system. These characteristics are subsequently translated into pharmacophore features for efficient high-throughput virtual screening. A retrospective evaluation revealed a successful 3D pharmacophore generation for three out of five pharmaceutically relevant targets. Two targets changed the ligand bound conformation quickly after initiating the ligand-bound conformation ultimately resulting in the absence of known important features and misplaced exclusion volumes.

Since 3D pharmacophore generation with PyRod was successful in only one out of two GPCR test systems, we were curious if PyRod can generate predictive 3D pharmacophores for MCHR1 (see section 4.4). Similar to H_3R, MCHR1 contains an aspartate

at position 3.32. PyRod suggests this area to be favorable for placing positive ionizable features and associated hydrogen bond donor features. These shared characteristics indicates interactions with $D^{3.32}$ to be a prerequisite to identify dual modulators of H_3R and MCHR1. Additionally, areas close to the three flexible glutamines were predicted to be favorable for hydrogen bonding interactions highlighting their potential importance for ligand binding. In total, 15 pharmacophore features were found to be favorably arranged in the binding pocket and were highly rated according to the corresponding features scores reported by PyRod. A combinatorial processing of this super pharmacophore with subsequent retrospective validation revealed several 3D pharmacophores able to discriminate between known actives and decoys.

The best 3D pharmacophore according to the early enrichment factor ($EF_{1\%} = 27.1$) constitutes of a positive ionizable feature next to $D^{3.32}$ accompanied by a hydrophobic feature as well as an aromatic feature and a hydrogen bond acceptor next to the extracellular loops. However, this 3D pharmacophore only finds 6 % of the actives and thus shows low sensitivity towards diverse MCHR1 antagonists. In contrast, the 3D pharmacophore with the highest hit rate among actives (35 %) has its hydrophobic feature located next to the three flexible glutamines and does not contain the aromatic feature. However, the increased sensitivity for actives coincides with a decrease in early enrichment ($EF_{1\%} = 4.0$) indicating lower hit rates in prospective virtual screening. In comparison to ligand-based 3D pharmacophores evaluated on the same data sets, PyRod pharmacophores achieved a higher sensitivity which further underlines their usefulness.

Taken together, PyRod was successfully employed to generate highly predictive 3D pharmacophore models for MCHR1. The pharmacophore generation procedure is thereby not dependent on docking poses of known active ligands but on water dynamics from MD simulation. Furthermore, these findings highlight the usefulness of PyRod in virtual screening campaigns targeting protein binding sites with highly flexible side chains.

Conclusion

The continuously increasing prevalence of obesity and overweight reveals an unmet challenge for public health care systems. The complex etiology of obesity hampered the development of an effective pharmaceutical treatment. Multi-target drugs may overcome the limitation of traditional pharmaceutics by modulating multiple nodes in the biological network resulting in improved efficacy and reduced side effects. In this thesis, the multi-target concept was applied to identify targets for potential mulit-target drug development against obesity. Furthermore, novel computational tools were developed to rationally gain and exploit information about the interactions of ligands and their macromolecular target.

A systematic data mining approach employing freely available activity data of ligands modulating obesity relevant targets enabled the identification of several promising target pairs that could be exploited for the development of multi-target drugs against obesity. The most promising target pair, H_3R and MCHR1, was further characterized, since we found evidence for a possible synergistic effect when antagonizing both receptors simultaneously. A virtual screening campaign assessing the 3D similarity of molecules targeting the H_3R and MCHR1 led to the identification of three molecules with prior unknown multi-target character.

Understanding the molecular interactions formed between active ligands and their macromolecular target is an integral part of modern drug design. However, structural data was not available for H_3R and MCHR1 at the time of this thesis. Hence, atomistic models were built for both receptors by homology modeling. Although frequently successful, homology models hold the risk of modeling errors and similar to x-ray structures only represent a single conformation of a dynamic system. Thus two novel methods were developed to support drug design campaigns relying on homology models.

The first method exploits a charged interactions towards $D^{3.32}$ that is known to be essential for ligand binding at the orthosteric binding pocket of aminergic receptors. The analysis of docking poses of a distinct ligand series at 1000 homology models of H_3R resulted in the identification of a side chain conformation of $E^{5.46}$ crucial for appropriate ligand placement. A subsequent docking study followed by 3D pharmacophore screening yielded two novel ligands of H_3R with nanomolar affinity. However,

this approach could not be applied to MCHR1, since required data about essential interactions between ligands and receptor is currently not available.

The second method was developed to escape the dependence of structure-based virtual screening campaigns on interactions formed between ligands and their molecular target. PyRod analyzes the protein environment of water molecules in MD simulations and allows the generation of highly efficient 3D pharmacophore models. Using this method, we generated 3D pharmacophore models for MCHR1 that are able to discriminate active molecules from decoys. This can be extremely useful, since confirmed hits deliver a potential binding hypothesis that can be exploited in subsequent lead optimizations. Also, PyRod is not limited to this receptor and can be applied to other macromolecular targets.

Taken together, this thesis reports the identification of H_3R and MCHR1 as promising target pair for the development of multi-target drugs against obesity. Three novel molecules were identified with prior unknown nanomolar affinity to both receptors. Additionally, two novel methods were developed that can support homology modeling studies and allow for generation of 3D pharmacophores from MD simulations.

Bibliography

1. Banerjee, S. *et al.* Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *New England Journal of Medicine* **377,** 13–27 (2017).

2. Heymsfield, S. B. *et al.* Mechanisms, Pathophysiology, and Management of Obesity. *New England Journal of Medicine* **376,** 254–266 (2017).

3. *World Health Organization, accessed: 2019-07-27* https://www.who.int/gho/ncd/risk_factors/overweight_obesity/obesity_adults/en/.

4. Kim, K.-S. *et al.* Signalling From the Periphery to the Brain That Regulates Energy Homeostasis. *Nature Reviews Neuroscience* **19,** 185–196 (2018).

5. Srivastava, G. *et al.* Current Pharmacotherapy for Obesity. *Nature Reviews Endocrinology* **14,** 12–24 (2018).

6. Chakrabarti, R. Pharmacotherapy of Obesity: Emerging Drugs and Targets. *Expert Opinion on Therapeutic Targets* **13,** 195–207 (2009).

7. Schaller, D. *et al.* Systematic Data Mining Reveals Synergistic H3R/MCHR1 Ligands. *ACS Medicinal Chemistry Letters* **8,** 648–653 (2017).

8. Gaulton, A. *et al.* The ChEMBL Database in 2017. *Nucleic Acids Research* **45,** D945–D954 (2017).

9. Heisler, L. K. *et al.* Serotonin Reciprocally Regulates Melanocortin Neurons to Modulate Food Intake. *Neuron* **51,** 239–49 (2006).

10. Boutant, M. *et al.* SIRT1 Enhances Glucose Tolerance by Potentiating Brown Adipose Tissue Function. *Molecular Metabolism* **4,** 118–131 (2015).

11. Gross, B. *et al.* PPARs in Obesity-Induced T2DM, Dyslipidaemia and NAFLD. *Nature Reviews Endocrinology* **13,** 36–49 (2017).

12. Roth, J. D. Amylin and the Regulation of Appetite and Adiposity: Recent Advances in Receptor Signaling, Neurobiology and Pharmacology. *Current Opinion in Endocrinology, Diabetes and Obesity* **20,** 8–13 (2013).

13. Collins, S. *et al.* The Beta-Adrenergic Receptors and the Control of Adipose Tissue Metabolism and Thermogenesis. *Recent Progress in Hormone Research* **56,** 309–28 (2001).

14. Ramos-Álvarez, I. *et al.* Bombesin Receptor Subtype-3 (BRS-3), a Novel Candidate as Therapeutic Molecular Target in Obesity and Diabetes. *Molecular and Cellular Endocrinology* **367,** 109–115 (2013).

15. Perry, B. *et al.* Appetite Regulation and Weight Control: The Role of Gut Hormones. *Nutrition & Diabetes* **2,** e26 (2012).

16. Grover, G. J. *et al.* Therapeutic Potential for Thyroid Hormone Receptor-Beta Selective Agonists for Treating Obesity, Hyperlipidemia and Diabetes. *Current Vascular Pharmacology* **5,** 141–54 (2007).

17. Warne, J. P. *et al.* Metabolic Transceivers: In Tune With the Central Melanocortin System. *Trends in Endocrinology and Metabolism* **24,** 68–75 (2013).

18. Ishihara, A. *et al.* Neuropeptide Y Receptors as Targets of Obesity Treatment. *Expert Opinion on Therapeutic Patents* **16,** 1701–1712 (2006).

19. Xu, T.-R. *et al.* Orexin Receptors: Multi-Functional Therapeutic Targets for Sleeping Disorders, Eating Disorders, Drug Addiction, Cancers and Other Physiological Disorders. *Cellular Signalling* **25,** 2413–23 (2013).

20. Heal, D. J. *et al.* Selective 5-HT6 Receptor Ligands: Progress in the Development of a Novel Pharmacological Approach to the Treatment of Obesity and Related Metabolic Disorders. *Pharmacology & Therapeutics* **117,** 207–31 (2008).

21. Watkins, B. A. *et al.* The Endocannabinoid System: Directing Eating Behavior and Macronutrient Metabolism. *Frontiers in Psychology* **5,** 1506 (2014).

22. Chapman, K. *et al.* 11B-Hydroxysteroid Dehydrogenases: Intracellular Gate-Keepers of Tissue Glucocorticoid Action. *Physiological Reviews* **93,** 1139–206 (2013).

23. Mastorakos, G. *et al.* The Hypothalamic-Pituitary-Adrenal Axis in the Neuroendocrine Regulation of Food Intake and Obesity: The Role of Corticotropin Releasing Hormone. *Nutritional Neuroscience* **7,** 271–80 (2004).

24. Fang, P. *et al.* Galanin and Its Receptors: A Novel Strategy for Appetite Control and Obesity Therapy. *Peptides* **36,** 331–339 (2012).

25. Strable, M. S. *et al.* Genetic Control of De Novo Lipogenesis: Role in Diet-Induced Obesity. *Critical Reviews in Biochemistry and Molecular Biology* **45,** 199–214 (2010).

26. Masaki, T. *et al.* Therapeutic Approach of Histamine H3 Receptors in Obesity. *Recent Patents on CNS Drug Discovery* **2,** 238–40 (2007).

27. Borowsky, B. *et al.* Antidepressant, Anxiolytic and Anorectic Effects of a Melanin-Concentrating Hormone-1 Receptor Antagonist. *Nature Medicine* **8,** 825–830 (2002).

28. Karlsson, H. K. *et al.* Obesity Is Associated with Decreased μ-Opioid But Unaltered Dopamine D2 Receptor Availability in the Brain. *Journal of Neuroscience* **35,** 3959–3965 (2015).

29. Bruce, C. R. *et al.* Overexpression of Carnitine Palmitoyltransferase-1 in Skeletal Muscle is Sufficient to Enhance Fatty Acid Oxidation and Improve High-Fat Diet-Induced Insulin Resistance. *Diabetes* **58,** 550–8 (2009).

30. Yen, C.-L. E. *et al.* Thematic Review Series: Glycerolipids. DGAT Enzymes and Triacylglycerol Biosynthesis. *Journal of Lipid Research* **49,** 2283–301 (2008).

31. Lodhi, I. J. *et al.* Inhibiting Adipose Tissue Lipogenesis Reprograms Thermogenesis and PPARγ Activation to Decrease Diet-Induced Obesity. *Cell Metabolism* **16,** 189–201 (2012).

32. Hvizdos, K. M. *et al.* Orlistat: A Review of Its Use in the Management of Obesity. *Drugs* **58,** 743–60 (1999).

33. Allison, M. B. *et al.* 20 Years of Leptin: Connecting Leptin Signaling to Biological Function. *The Journal of Endocrinology* **223,** T25–T35 (2014).

34. Sampath, H. *et al.* Role of Stearoyl-CoA Desaturase-1 in Skin Integrity and Whole Body Energy Balance. *Journal of Biological Chemistry* **289,** 2482–2488 (2014).

35. Carow, B. *et al.* SOCS3, a Major Regulator of Infection and Inflammation. *Frontiers in Immunology* **5,** 58 (2014).

36. Picardi, P. K. *et al.* Reduction of Hypothalamic Protein Tyrosine Phosphatase Improves Insulin and Leptin Resistance in Diet-Induced Obese Rats. *Endocrinology* **149,** 3870–80 (2008).

37. Sawa, M. *et al.* Recent Developments in the Design of Orally Bioavailable β3-Adrenergic Receptor Agonists. *Current Medicinal Chemistry* **13,** 25–37 (2006).

38. Pi-Sunyer, X. *et al.* C-Terminal Octapeptide of Cholecystokinin Decreases Food Intake in Obese Men. *Physiology & Behavior* **29,** 627–630 (1982).

39. Beglinger, C. *et al.* Loxiglumide, a CCK-A Receptor Antagonist, Stimulates Calorie Intake and Hunger Feelings in Humans. *American Journal of Physiology* **280,** R1149–R1154 (2001).

40. Staels, B. *et al.* Fibrates and Future PPARα Agonists in the Treatment of Cardiovascular Disease. *Nature Clinical Practice Cardiovascular Medicine* **5,** 542–553 (2008).

41. Tanaka, T. *et al.* Activation of Peroxisome Proliferator-Activated Receptor Induces Fatty Acid-Oxidation in Skeletal Muscle and Attenuates Metabolic Syndrome. *Proceedings of the National Academy of Sciences* **100,** 15924–15929 (2003).

42. Petrovic, N. *et al.* Chronic Peroxisome Proliferator-Activated Receptor γ Activation of Epididymally Derived White Adipocyte Cultures Reveals a Population of Thermogenically Competent, UCP1-Containing Adipocytes Molecularly Distinct from Classic Brown Adipocytes. *Journal of Biological Chemistry* **285,** 7153–7164 (2010).

43. Soccio, R. E. *et al.* Thiazolidinediones and the Promise of Insulin Sensitization in Type 2 Diabetes. *Cell Metabolism* **20,** 573–591 (2014).

44. Wing, R. R. *et al.* Benefits of Modest Weight Loss in Improving Cardiovascular Risk Factors in Overweight and Obese Individuals With Type 2 Diabetes. *Diabetes Care* **34,** 1481–1486 (2011).

45. Anderson, J. W. *et al.* Long-Term Weight-Loss Maintenance: A Meta-Analysis of US Studies. *The American Journal of Clinical Nutrition* **74,** 579–584 (2001).

46. Sjöström, L. *et al.* Effects of Bariatric Surgery on Mortality in Swedish Obese Subjects. *New England Journal of Medicine* **357,** 741–752 (2007).

47. Sjöström, L. *et al.* Bariatric Surgery and Long-Term Cardiovascular Events. *The Journal of the American Medical Association* **307,** 56 (2012).

48. Kashyap, S. R. *et al.* Metabolic Effects of Bariatric Surgery in Patients With Moderate Obesity and Type 2 Diabetes: Analysis of a Randomized Control Trial Comparing Surgery With Intensive Medical Treatment. *Diabetes Care* **36,** 2175–2182 (2013).

49. Hutter, M. M. *et al.* First Report From the American College of Surgeons Bariatric Surgery Center Network. *Annals of Surgery* **254,** 410–422 (2011).

50. Yanovski, S. Z. *et al.* Long-Term Drug Treatment for Obesity: A Systematic and Clinical Review. *The Journal of the American Medical Association* **311,** 74–86 (2014).

51. Aronne, L. J. *et al.* Evaluation of Phentermine and Topiramate Versus Phentermine/Topiramate Extended-Release in Obese Adults. *Obesity* **21,** 2163–2171 (2013).

52. Alfaris, N. *et al.* Combination Phentermine and Topiramate Extended Release in the Management of Obesity. *Expert Opinion on Pharmacotherapy* **16,** 1263–1274 (2015).

53. Thomsen, W. J. *et al.* Lorcaserin, a Novel Selective Human 5-Hydroxy tryptamine2C Agonist: In Vitro and in Vivo Pharmacological Characterization. *Journal of Pharmacology and Experimental Therapeutics* **325,** 577–587 (2008).

54. Billes, S. K. *et al.* Naltrexone/Bupropion for Obesity: An Investigational Combination Pharmacotherapy for Weight Loss. *Pharmacological Research* **84,** 1–11 (2014).

55. Astrup, A. *et al.* Safety, Tolerability and Sustained Weight Loss Over 2 Years With the Once-Daily Human GLP-1 Analog, Liraglutide. *International Journal of Obesity* **36,** 843–854 (2012).

56. Hauser, A. S. *et al.* Trends in GPCR Drug Discovery: New Agents, Targets and Indications. *Nature Reviews Drug Discovery* **16,** 829–842 (2017).

57. Sloop, K. W. *et al.* The Current State of GPCR-Based Drug Discovery to Treat Metabolic Disease. *British Journal of Pharmacology* **175,** 4060–4071 (2018).

58. Venkatakrishnan, A. J. *et al.* Molecular Signatures of G-Protein-Coupled Receptors. *Nature* **494,** 185–194 (2013).

59. Ballesteros, J. A. *et al.* Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors. *Methods in Neurosciences* **25,** 366–428 (1995).

60. Isberg, V. *et al.* Generic GPCR Residue Numbers - Aligning Topology Maps While Minding the Gaps. *Trends in Pharmacological Sciences* **36,** 22–31 (2015).

61. Michino, M. *et al.* Structural Basis for Na+ - Sensitivity in Dopamine D2 and D3 Receptors. *Chemical Communications* **51,** 8618–8621 (2015).

62. Hua, T. *et al.* Crystal Structure of the Human Cannabinoid Receptor CB1. *Cell* **167,** 750–762 (2016).

63. Isberg, V. *et al.* GPCRdb: An Information System for G Protein-Coupled Receptors. *Nucleic Acids Research* **44,** D356–D364 (2016).

64. Panula, P. *et al.* International Union of Basic and Clinical Pharmacology. Histamine Receptors. *Pharmacological Reviews* **67,** 601–655 (2015).

65. Nieto-Alamilla, G. *et al.* The Histamine H3 Receptor: Structure, Pharmacology, and Function. *Molecular Pharmacology* **90,** 649–673 (2016).

66. Syed, Y. Y. Pitolisant: First Global Approval. *Drugs* **76,** 1313–1318 (2016).

67. Ookuma, K. *et al.* Neuronal Histamine in the Hypothalamus Suppresses Food Intake in Rats. *Brain Research* **628,** 235–242 (1993).

68. Orthen-Gambill, N. *et al.* FMH-Induced Decrease in Central Histamine Levels Produces Increased Feeding and Body Weight in Rats. *Physiology & Behavior* **51,** 891–893 (1992).

69. Kotanska, M. *et al.* The Histamine H3 Receptor Inverse Agonist Pitolisant Reduces Body Weight in Obese Mice. *Naunyn-Schmiedeberg's Archives of Pharmacology* **391,** 875–881 (2018).

70. Hancock, A. A. *et al.* Antiobesity Effects of A-331440, a Novel Non-Imidazole Histamine H3 Receptor Antagonist. *European Journal of Pharmacology* **487,** 183–197 (2004).

71. Malmlöf, K. *et al.* Influence of a Selective Histamine H3 Receptor Antagonist on Hypothalamic Neural Activity, Food Intake and Body Weight. *International Journal of Obesity* **29,** 1402–1412 (2005).

72. Hancock, A. a. *et al.* Assessment of Pharmacology and Potential Anti-Obesity Properties of H3 Receptor Antagonists/Inverse Agonists. *Expert Opinion on Investigational Drugs* **14,** 223–241 (2005).

73. De Lera Ruiz, M. *et al.* Bicyclic and Tricyclic Heterocycle Derivatives as Histamine H3 Receptor Antagonists for the Treatment of Obesity. *Bioorganic & Medicinal Chemistry Letters* **23,** 6004–9 (2013).

74. Ghamari, N. *et al.* Histamine H3 Receptor Antagonists/Inverse Agonists: Where Do They Go? *Pharmacology & Therapeutics* **200,** 69–84 (2019).

75. MacNeil, D. J. The Role of Melanin-Concentrating Hormone and Its Receptors in Energy Homeostasis. *Frontiers in Endocrinology* **4,** 1–14 (2013).

76. Arrigoni, E. *et al.* To Eat or to Sleep: That Is a Lateral Hypothalamic Question. *Neuropharmacology* **154,** 34–49 (2019).

77. Pissios, P. *et al.* Melanin-Concentrating Hormone Receptor 1 Activates Extracellular Signal-Regulated Kinase and Synergizes with Gs-Coupled Pathways. *Endocrinology* **144,** 3514–3523 (2003).

78. Sailer, A. W. *et al.* Identification and Characterization of a Second Melanin-Concentrating Hormone Receptor, MCH-2R. *Proceedings of the National Academy of Sciences* **98,** 7564–7569 (2001).

79. Zhou, D. *et al.* Enhanced Running Wheel Activity of Both Mch1r- and Pmch-Deficient Mice. *Regulatory Peptides* **124,** 53–63 (2005).

80. Willie, J. T. *et al.* Abnormal Response of Melanin-Concentrating Hormone Deficient Mice to Fasting: Hyperactivity and Rapid Eye Movement Sleep Suppression. *Neuroscience* **156,** 819–829 (2008).

81. Ludwig, D. S. *et al.* Melanin-Concentrating Hormone Overexpression in Transgenic Mice Leads to Obesity and Insulin Resistance. *Journal of Clinical Investigation* **107,** 379–386 (2001).

82. Alon, T. *et al.* Late-Onset Leanness in Mice With Targeted Ablation of Melanin Concentrating Hormone Neurons. *The Journal of Neuroscience* **26,** 389–397 (2006).

83. Tan, C. P. *et al.* Melanin-Concentrating Hormone Receptor Subtypes 1 and 2: Species-Specific Gene Expression. *Genomics* **79,** 785–792 (2002).

84. Chee, M. J. S. *et al.* Expression of Melanin-Concentrating Hormone Receptor 2 Protects Against Diet-Induced Obesity in Male Mice. *Endocrinology* **155,** 81–88 (2014).

85. Hertzog, D. L. *et al.* The Discovery and Optimization of Pyrimidinone-Containing MCHR1 Antagonists. *Bioorganic & Medicinal Chemistry Letters* **16,** 4723–4727 (2006).

86. Motani, A. S. *et al.* Evaluation of AMG 076, a Potent and Selective MCHR1 Antagonist, in Rodent and Primate Obesity Models. *Pharmacology Research & Perspectives* **1,** e00003 (2013).

87. Moore, N. A. *et al.* From Preclinical to Clinical Development: The Example of a Novel Treatment for Obesity. *Neurobiology of Disease* **61,** 47–54 (2014).

88. Devasthale, P. *et al.* Non-Basic Azolotriazinone MCHR1 Antagonists for the Treatment of Obesity: An Empirical Brain-Exposures-Driven Candidate Selection for in Vivo Efficacy Studies. *Bioorganic & Medicinal Chemistry Letters* **25,** 4412–4418 (2015).

89. Johansson, A. *et al.* Discovery of (3-(4-(2-Oxa-6-azaspiro[3.3]heptan-6-ylmethyl) phenoxy)azetidin-1-yl) (5-(4-methoxyphenyl)-1,3,4-oxadiazol-2 -yl) methanone (AZD1979), a Melanin Concentrating Hormone Receptor 1 (MCHr1) Antagonist with Favorable Physicochemical Properties. *Journal of Medicinal Chemistry* **59,** 2497–2511 (2016).

90. Johansson, A. Evolution of Physicochemical Properties of Melanin Concentrating Hormone Receptor 1 (MCHr1) Antagonists. *Bioorganic and Medicinal Chemistry Letters* **26,** 4559–4564 (2016).

91. Della-Zuana, O. *et al.* Peripheral Injections of Melanin-Concentrating Hormone Receptor 1 Antagonist S38151 Decrease Food Intake and Body Weight in Rodent Obesity Models. *Frontiers in Endocrinology* **3** (2012).

92. Agis-Torres, A. *et al.* Multi-Target-Directed Ligands and Other Therapeutic Strategies in the Search of a Real Solution for Alzheimer's Disease. *Current Neuropharmacology* **12,** 2–36 (2014).

93. Anighoro, A. *et al.* Polypharmacology: Challenges and Opportunities in Drug Discovery. *Journal of Medicinal Chemistry* **57,** 7874–7887 (2014).

94. Mestres, J. *et al.* Data Completeness-The Achilles Heel of Drug-Target Networks. *Nature Biotechnology* **26,** 983–984 (2008).

95. Csermely, P. *et al.* Structure and Dynamics of Molecular Networks: A Novel Paradigm of Drug Discovery. *Pharmacology & Therapeutics* **138,** 333–408 (2013).

96. Roth, B. L. *et al.* Magic Shotguns Versus Magic Bullets: Selectively Non-Selective Drugs for Mood Disorders and Schizophrenia. *Nature Reviews Drug Discovery* **3,** 353–359 (2004).

97. Sliwoski, G. *et al.* Computational Methods in Drug Discovery. *Pharmacological Reviews* **66,** 334–395 (2014).

98. Maggiora, G. *et al.* Molecular Similarity in Medicinal Chemistry. *Journal of Medicinal Chemistry* **57,** 3186–3204 (2014).

99. Lipinski, C. A. *et al.* Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Advanced Drug Delivery Reviews* **46,** 3–26 (2001).

100. D. Shultz, M. Two Decades Under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs. *Journal of Medicinal Chemistry* **62,** 1701–1714 (2018).

101. Willett, P. in *Chemoinformatics and Computational Chemical Biology* 133–158 (Humana Press, Totowa, 2010).

102. L. Durant, J. *et al.* Reoptimization of MDL Keys for Use in Drug Discovery. *Journal of Chemical Information and Computer Sciences* **42,** 1273–1280 (2002).

103. Rogers, D. *et al.* Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* **50,** 742–754 (2010).

104. Bajusz, D. *et al.* Why is Tanimoto Index an Appropriate Choice for Finger-Print-Based Similarity Calculations? *Journal of Cheminformatics* **7,** 1–13 (2015).

105. Keiser, M. J. *et al.* Relating Protein Pharmacology by Ligand Chemistry. *Nature Biotechnology* **25,** 197–206 (2007).

106. Griffen, E. *et al.* Matched Molecular Pairs as a Medicinal Chemistry Tool. *Journal of Medicinal Chemistry* **54,** 7739–7750 (2011).

107. Venkatraman, V. *et al.* Comprehensive Comparison of Ligand-Based Virtual Screening Tools Against the DUD Data Set Reveals Limitations of Current 3D Methods. *Journal of Chemical Information and Modeling* **50,** 2079–2093 (2010).

108. Hawkins, P. C. D. *et al.* Comparison of Shape-Matching and Docking as Virtual Screening Tools. *Journal of Medicinal Chemistry* **50,** 74–82 (2007).

109. Burley, S. K. *et al.* RCSB Protein Data Bank: Biological Macromolecular Structures Enabling Research and Education in Fundamental Biology, Biomedicine, Biotechnology and Energy. *Nucleic Acids Research* **47,** D464–D474 (2019).

110. Berman, H. M. The Protein Data Bank. *Nucleic Acids Research* **28,** 235–242 (2000).

111. Zhao, Q. *et al.* Ice Breaking in GPCR Structural Biology. *Acta Pharmacologica Sinica* **33,** 324–334 (2012).

112. Schmidt, T. *et al.* Modelling Three-Dimensional Protein Structures for Applications in Drug Design. *Drug Discovery Today* **19,** 890–897 (2014).

113. Webb, B. *et al.* in *Current Protocols in Bioinformatics* 1–5 (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2016).

114.   Chemical Computing Group Inc. *Molecular Operating Environment* Montreal, QC, Canada, 2018.

115.   Altschul, S. F. *et al.* Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215,** 403–410 (1990).

116.   Needleman, S. B. *et al.* A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *Journal of Molecular Biology* **48,** 443–453 (1970).

117.   Rodriguez, R. *et al.* Homology Modeling, Model and Software Evaluation: Three Related Resources. *Bioinformatics* **14,** 523–528 (1998).

118.   Pagadala, N. S. *et al.* Software for Molecular Docking: A Review. *Biophysical Reviews* **9,** 91–102 (2017).

119.   Cole, J. *et al.* in *Virtual Screening in Drug Discovery* 379–415 (Taylor & Francis CRC Press, Boca Raton, 2005).

120.   Goodsell, D. S. *et al.* Automated Docking of Flexible Ligands: Applications of Autodock. *Journal of Molecular Recognition* **9,** 1–5 (1996).

121.   Rarey, M. *et al.* A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *Journal of Molecular Biology* **261,** 470–489 (1996).

122.   Friesner, R. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **47,** 1739–1749 (2004).

123.   Warren, G. L. *et al.* A Critical Assessment of Docking Programs and Scoring Functions. *Journal of Medicinal Chemistry* **49,** 5912–5931 (2006).

124.   Wang, Z. *et al.* Comprehensive Evaluation of Ten Docking Programs on a Diverse Set of Protein-Ligand Complexes: The Prediction Accuracy of Sampling Power and Scoring Power. *Physical Chemistry Chemical Physics* **18,** 12964–12975 (2016).

125.   Reymond, J.-L. *et al.* Chemical Space as a Source for New Drugs. *MedChemComm* **1,** 30 (2010).

126.   Gromski, P. S. *et al.* *How to Explore Chemical Space Using Algorithms and Automation* 2019. http://www.nature.com/articles/s41570-018-0066-y.

127.   MacArron, R. *et al.* Impact of High-Throughput Screening in Biomedical Research. *Nature Reviews Drug Discovery* **10,** 188–195 (2011).

128.   Seidel, T. *et al.* Strategies for 3D Pharmacophore-Based Virtual Screening. *Drug Discovery Today: Technologies* **7,** e221–e228 (2010).

129.   Vuorinen, A. *et al.* Methods for Generating and Applying Pharmacophore Models as Virtual Screening Filters and for Bioactivity Profiling. *Methods* **71,** 113–134 (2015).

130. Mortier, J. *et al.* The Impact of Molecular Dynamics on Drug Design: Applications for the Characterization of Ligand-Macromolecule Complexes. *Drug Discovery Today* **20,** 686–702 (2015).

131. De Vivo, M. *et al.* Role of Molecular Dynamics and Related Methods in Drug Discovery. *Journal of Medicinal Chemistry* **59,** 4035–4061 (2016).

132. Bowers, K. *et al.* Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *ACM/IEEE SC 2006 Conference (SC'06)* (2006).

133. Salomon-Ferrer, R. *et al.* An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **3,** 198–210 (2013).

134. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* **30,** 1545–1614 (2009).

135. Berendsen, H. *et al.* GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications* **91,** 43–56 (1995).

136. Riniker, S. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *Journal of Chemical Information and Modeling* **58,** 565–578 (2018).

137. Segala, E. *et al.* Controlling the Dissociation of Ligands From the Adenosine A 2A Receptor Through Modulation of Salt Bridge Strength. *Journal of Medicinal Chemistry* **59,** 6470–6479 (2016).

138. Triballeau, N. *et al.* Virtual Screening Workflow Development Guided by the Receiver Operating Characteristic Curve Approach. Application to High Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *Journal of Medicinal Chemistry* **48,** 2534–47 (2005).

139. Kirchmair, J. *et al.* Evaluation of the Performance of 3D Virtual Screening Protocols: RMSD Comparisons, Enrichment Assessments, and Decoy Selection- What Can we Learn From Earlier Mistakes? *Journal of Computer-Aided Molecular Design* **22,** 213–228 (2008).

140. Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets. *Journal of Chemical Information and Computer Sciences* **39,** 747–750 (1999).

141. Cirauqui, N. *et al.* Building a MCHR1 Homology Model Provides Insight Into the Receptor-Antagonist Contacts That Are Important for the Development of New Anti-Obesity Agents. *Bioorganic and Medicinal Chemistry* **18,** 7365–7379 (2010).

142. Hopkins, A. L. *et al.* The Role of Ligand Efficiency Metrics in Drug Discovery. *Nature Reviews Drug Discovery* **13,** 105–121 (2014).

143. Parks, G. S. *et al.* Histamine Inhibits the Melanin-Concentrating Hormone System: Implications for Sleep and Arousal. *The Journal of Physiology* **592,** 2183–96 (2014).

144. Kufareva, I. *et al.* Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. *Structure* **22,** 1120–1139 (2014).

145. Miao, Z. *et al.* Quantifying Side-Chain Conformational Variations in Protein Structure. *Scientific Reports* **6,** 37024 (2016).

146. Michino, M. *et al.* What Can Crystal Structures of Aminergic Receptors Tell us About Designing Subtype-Selective Ligands? *Pharmacological Reviews* **67,** 198–213 (2015).

147. Nikolic, K. *et al.* Pharmacophore Modeling, Drug Design and Virtual Screening on Multi-Targeting Procognitive Agents Approaching Histaminergic Pathways. *Journal of the Taiwan Institute of Chemical Engineers* **46,** 15–29 (2015).

148. Uveges, A. J. *et al.* The Role of Transmembrane Helix 5 in Agonist Binding to the Human H3 Receptor. *The Journal of Pharmacology and Experimental Therapeutics* **301,** 451–8 (2002).

149. Kiss, R. *et al.* Structure-Based Discovery and Binding Site Analysis of Histamine Receptor Ligands. *Expert Opinion on Drug Discovery* **11,** 1165–1185 (2016).

150. Kooistra, A. *et al.* A Structural Chemogenomics Analysis of Aminergic GPCRs: Lessons for Histamine Receptor Ligand Design. *British Journal of Pharmacology* **170,** 101–126 (2013).

151. Moguilevsky, N. *et al.* Pharmacological and Functional Characterisation of the Wild-Type and Site-Directed Mutants of the Human H1 Histamine Receptor Stably Expressed in CHO Cells. *Journal of Receptor and Signal Transduction Research* **15,** 91–102.

152. Bruysters, M. *et al.* Mutational Analysis of the Histamine H1-Receptor Binding Pocket of Histaprodifens. *European Journal of Pharmacology* **487,** 55–63 (2004).

List of Publications

1. Pach, S. *et al.* Catching a Moving Target: Comparative Modeling of Flaviviral NS2B-NS3 Reveals Small Molecule Zika Protease Inhibitors. *ACS Medicinal Chemistry Letters,* acsmedchemlett.9b00629 (2020).

2. Schaller, D. *et al.* Next Generation 3D Pharmacophore Modeling. *Wiley Interdisciplinary Reviews: Computational Molecular Science,* e1468 (2020).

3. Aygün Cevher, H. *et al.* Discovery of Michael Acceptor Containing 1,4-Dihydro-Pyridines as First Covalent Inhibitors of L-/T-Type Calcium Channels. *Bioorganic Chemistry* **91,** 103187 (2019).

4. Schaller, D. *et al.* Ligand-Guided Homology Modeling Drives Identification of Novel Histamine H3 Receptor Ligands. *PLOS ONE* **14,** e0218820 (2019).

5. Schaller, D. *et al.* PyRod: Tracing Water Molecules in Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **59,** 2818–2829 (2019).

6. Naß, A. *et al.* Assessment of Flexible Shape Complementarity: New Opportunities to Explain and Induce Selectivity in Ligands of Protein Tyrosine Phosphatase 1B. *Molecular Informatics* **38,** 1800141 (2019).

7. Schaller, D. *et al.* Binding Mechanism Investigations Guiding the Synthesis of Novel Condensed 1,4-Dihydropyridine Derivatives With L-/T-Type Calcium Channel Blocking Activity. *European Journal of Medicinal Chemistry* **155,** 1–12 (2018).

8. Schaller, D. *et al.* Systematic Data Mining Reveals Synergistic H3R/MCHR1 Ligands. *ACS Medicinal Chemistry Letters* **8,** 648–653 (2017).

Selbständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig und ausschließlich unter Zuhilfenahme der angegeben Literatur erstellt habe.

Berlin, den 29.06.2020

David Schaller