Edinburgh Research Explorer

Design of drug-like hepsin inhibitors against prostate cancer and kidney stones

OPEN ACCESS
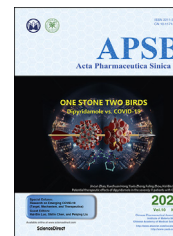
ORIGINAL ARTICLE

Design of drug-like hepsin inhibitors against prostate cancer and kidney stones

Vincent Blay[a,b,*], Mu-Chun Li[c], Sunita P. Ho[a,b], Mashall L. Stoller[a,b], Hsing-Pang Hsieh[c], Douglas R. Houston[d]

[a]Division of Biomaterials and Bioengineering, School of Dentistry, University of California San Francisco, San Francisco, CA 94143, USA
[b]Department of Urology, School of Medicine, University of California San Francisco, San Francisco, CA 94143, USA
[c]Institute of Biotechnology and Pharmaceutical Research, National Health Research Institutes, Zhunan, Taiwan 350, China
[d]University of Edinburgh, Institute of Quantitative Biology, Biochemistry and Biotechnology, Edinburgh, Scotland, EH9 3BF, UK

Abstract   Hepsin, a transmembrane serine protease abundant in renal endothelial cells, is a promising therapeutic target against several cancers, particularly prostate cancer. It is involved in the release and polymerization of uromodulin in the urine, which plays a role in kidney stone formation. In this work, we design new potential hepsin inhibitors for high activity, improved specificity towards hepsin, and promising ADMET properties. The ligands were developed *in silico* through a novel hierarchical pipeline. This pipeline explicitly accounts for off-target binding to the related serine proteases matriptase and HGFA (human hepatocyte growth factor activator). We completed the pipeline incorporating ADMET properties of the candidate inhibitors into custom multi-objective optimization functions. The ligands designed show excellent prospects for targeting hepsin *via* the blood stream and the urine and thus enable key experimental studies. The computational pipeline proposed is remarkably cost-efficient and can be easily adapted for designing inhibitors against new drug targets.

*Corresponding author. Tel.: +1 415 5142818.
E-mail address: vincent.blayroger@ucsf.edu (Vincent Blay).

1. Introduction

Kidney stones affect 8.8% of the current world population[1]. They affect all geographical regions and all age groups and impact the economy and welfare of society[2]. Kidney stones are a global health problem.

In general, kidney stones are composed of mixed calcium oxalates (mainly calcium oxalate monohydrate) and calcium phosphate (80% of the cases), struvite or magnesium ammonium phosphate (10% of the cases, more prevalent in women), urate (7%), cystine (2%, due to a genetic disorder), and lithogenic drugs (1%)[3]. Although the urine of humans is normally supersaturated with respect to calcium and oxalate[4], the deposition of crystals in most individuals is prevented during the residence time of urine in the body. This is achieved through a kinetic control of the mineralization reaction by inhibitory molecules naturally present in the urine. Some of these inhibitors are ions and proteins coming from blood filtration, such as magnesium, citrate, bikunin and prothrombin, whereas others are secreted by cells in the nephron and other parts of the kidney, such as osteopontin, nephrocalcin, and uromodulin (Umod)[3,5-7]. The mechanism of kidney stone formation is typically divided into nucleation, single crystal growth, crystal aggregation, and retention of crystal aggregates, often on the distal region of the kidney papilla[3,6,7]. Various inhibitors prevent mineralization at the nucleation stage. For instance, citrate chelates calcium ions in solution and magnesium binds to oxalate anions in urine, reducing their activity. Nonetheless, other inhibitors act at later stages of the mineral formation process and are effective since small crystals can be excreted during micturition.

Mutations in inhibitors of urine mineralization can decrease their preventive activity and increase the risk of stone formation[8-10]. Besides, nucleation of supersaturated solutions is often catalyzed by the presence of solid surfaces. Membrane phospholipids, bacteria, and cell debris resulting from injury (due to excessive oxalate concentration, for instance) may promote the nucleation of kidney stones[3,11]. A good match between the atomic structures of the substrate and the mineral phase being deposited on it (i.e., a high interfacial correlation factor) favors this catalysis[12]. For instance, secondary nucleation of calcium oxalate may be facilitated by initial crystals of calcium phosphate, like those in Randall's plaque[13-15]. Likewise, some biomolecules in urine may act as promoters of mineralization, including osteopontin and uromodulin. Overall, it is thought that an imbalance between promoters and inhibitors of urine crystallization primes kidney stone formation[6,15].

Uromodulin, also known as uromucoid or Tamm-Horsfall protein (THP), is the most abundant protein in the urine of healthy individuals. It is produced by renal epithelial cells that line the thick ascending limb (TAL) of Henle's loop. Uromodulin contains an external hydrophobic patch (EHP) that is cleaved off before its release into urine (shedding), where it polymerizes into high-weight polymers via its zona pellucida (ZP) domain[16,17]. The biological role of uromodulin is still not fully understood. In healthy individuals, uromodulin helps to prevent crystal aggregation and bacterial infection and is involved in water and electrolyte homeostasis[18-22]. However, the role of uromodulin in crystal aggregation is affected by its glycosylation state (Umod has eight possible N-glycosylation sites), as well as by the pH and ionic strength of urine[23-25]. Uromodulin from stone formers and that found in the matrix of kidney stones often present low levels of sialylation[23]. At urinary salt levels, desialylated Umod shows a high isoelectric point, which favors its aggregation and binding to other proteins in urine[23]. Thus, the polymerization of urinary Umod may contribute to pathologic crystal aggregation depending on its glycosylation state and the conditions of the urine in an individual. It also possible that, even in the same individual, membrane-bound uromodulin plays a different role on kidney stone progression, perhaps contributing to the anti-adherence of the TAL epithelium as it preserves the hydrophobic patch domain (Fig. 1). Recently, it was reported that hepsin (Hpn), a type II transmembrane serine protease (TTSP), catalyzes the cleavage of Umod in the apical membrane and promotes its secretion into urine[26]. Thus, it is important to study whether the specific inhibition of hepsin in certain stone formers, with the associated reduction in urinary uromodulin, could prevent or delay the formation of kidney stones in these patients.

Furthermore, hepsin overexpression been documented in several human cancers, including renal cell carcinoma, ovarian cancer, breast cancer, endometrial cancer, and, especially, prostate cancer[27-29]. In fact, hepsin overexpression is one of the most consistent biomarkers for malignant transformation from benign prostate hyperplasia[30]. Hepsin is thought to affect cancer cell migration, invasion and metastasis more than tumor growth[30]. Transgenic mice with high levels of hepsin in the prostate epithelium showed a more aggressive progression and metastasis by prostate cancer cells[31]. It has been proposed that the downregulation of hepsin inhibitors in epithelial cancers may be associated to desmosomal damage and loss of epithelial integrity[28]. At the same time, hepsin knockout mice have been shown to develop normally[32,33], suggesting that hepsin is not essential for normal cell growth or coagulation. Taken together, these results make hepsin a potential target for the treatment of several cancer types, and in fact a few inhibitors have been proposed recently to this end[28,34,35]. In vitro studies have shown that the growth of hepatoma cells was suppressed after treatment with anti-hepsin antibodies[36]. Hepsin inhibitors displayed anticancer properties in breast, prostate, colon, and lung cancer cells[28]. However, previous inhibitors exhibit significant affinity toward other trypsin-like serine proteases, such as matriptase and HGFA (hepatocyte growth factor activator). In particular, matriptase is also a type II transmembrane serine protease but, unlike other TTSPs, it is present in a wide range of tissues[37], hence the importance of considering it as a potential off-target when developing new hepsin inhibitors. Besides, previous inhibitors have been evaluated for activity in vitro with little or no consideration of their druglikeness or ADMET characteristics, which are particularly relevant for targeting hepsin in the nephron.

The development of improved hepsin inhibitors could enable future studies to advance our knowledge of the biological role of hepsin and uromodulin in the body. Based on our current understanding, it is conceivable that reducing the levels of polymerizable Umod in the urine of certain stone forming patients could have a therapeutic effect in kidney stone disease. Moreover, more specific small-molecule hepsin inhibitors could enable new chemotherapeutic agents for prostate cancer. In this work, we develop improved small-molecule competitive inhibitors of hepsin using an in-house computational pipeline. In addition to considering the affinity of the candidate molecules for hepsin, we explicitly considered their off-target effects on the related trypsin-like serine proteases matriptase and HGFA. We also estimated key ADMET properties of the ligands to optimize their availability in the urine. From a pharmacological point of view, this work is a particularly interesting case because of the target class selected
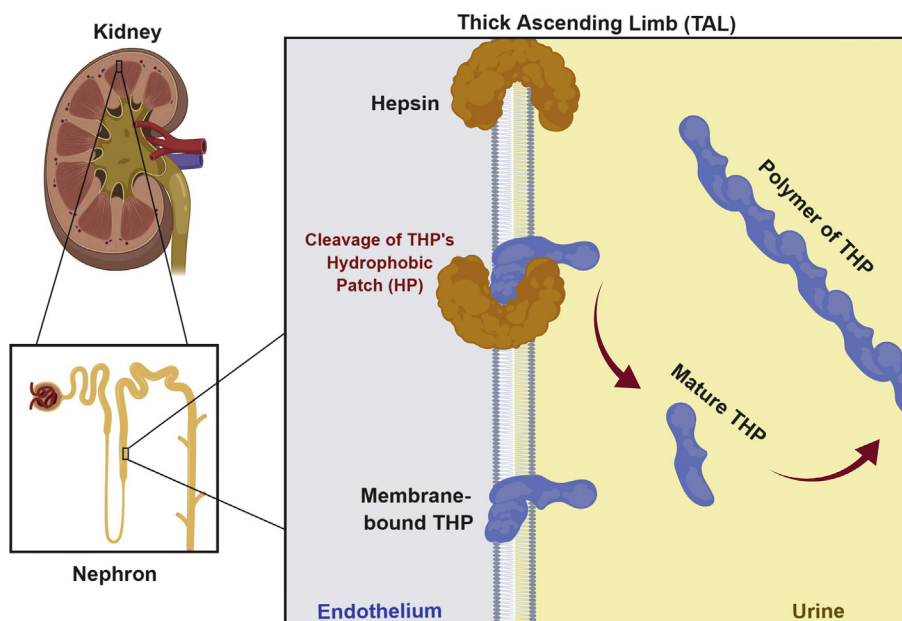
Figure 1 Scheme of uromodulin maturation and release into urine.

and the special requirements imposed on the drug candidates, which should achieve a relatively fast excretion from the circulatory system with minimal degradation and off-target binding.

2. Methods

2.1. Protein structures

The structure of hepsin in complex with 2-(2-hydroxyphenyl)-1*H*-benzimidazole-5-carboxamidine is available in the Protein Data Bank (entry 1P57)[38]. The high quality of the experimental data (resolution = 1.75 Å, overall completeness = 90.1%, overall R_{merge} = 4%) and of the modeling (R_{work} = 18.6%, R_{free} = 20.6%) are appropriate for the computational design of small molecules aiming to modulate its activity[39]. Moreover, the B-factors of the atoms in the protease active site are small and its fold shows a small root-mean-square deviation (RMSD) from that of related proteases in this region[38]. The structure of matriptase is available with PDB accession code 4O9V (resolution = 1.9 Å, R_{work} = 17.0%, R_{free} = 23.9%). The structure of HGFA is available with PDB accession code 2WUC (resolution = 2.7 Å, R_{work} = 22.6%, R_{free} = 27.4%).

2.2. Virtual screening

LIDAEUS (LIgand Discovery at Edinburgh UniverSity) is an in-house virtual screening software. In order to accelerate the screening procedure, a grid of sites is created in and around the protein active site, and these sites are classified as hydrogen bond donor (HBD), hydrogen bond acceptor (HBA) or hydrophobic depending on their chemical environment. An exhaustive fit of a given number of atoms from the rigid ligands onto the site points is carried out to identify energy-minimizing poses. The precision with which an atom is matched to a site, called "resolution" in this program, was set to 0.04 Å. The grids of sites employed for the three different targets are illustrated in Supporting Information Fig. S1 and provided as Supporting Information to this article

(files.mol). LIDAEUS uses a combination of two built-in scoring functions, a force field-based energy function and a knowledge-based pose interaction profile (PIP) function[40]. The "CAMEL-SICK2" library was virtually screened, which contains over 1.1 million ligands that passed a lead-like filter and contains multiple conformations of each ligand, totaling over 4.84×10^6 conformers screened[41].

2.3. Docking

AutoDock Vina, herein referred to as Vina, was used for the molecular docking of the most promising compounds. This software was developed at the Scripps Research Institute and is freely available on the Internet[42]. Vina uses a combination of empirical and knowledge-based scoring functions, which were trained using nonlinear regression on a very large dataset (PDBbind). Vina also implements an efficient search and optimization algorithm, the Iterated Local Search global optimizer[42]. This optimizer can be orders of magnitude faster than that in AutoDock 4.2, another docking software used in this work that relies on a Lamarckian genetic algorithm-based optimizer. Besides, Vina has built-in support for internal parallelism in its computations. We also used AutoDock 4.2, herein referred to as Vina, to independently evaluate a subset of the ligands. The free-energy scoring function of AutoDock is based on a linear regression analysis, the AMBER force field, and a set of protein-ligand complexes with known inhibition constants[43]. Docking computations were run on a 32-core computer, with external parallelism enabled by a custom-made script. In the case of Vina, an internal parallelism level of two was typically used[44]. Binding constants are reported at standard temperature (298 K).

2.4. ADMET profile

Optibrium StarDrop 6.5.1 was used to predict important ADMET properties of putative hepsin inhibitors. These include log*S*, log*P*, log*D*, 2C9 pK_i, hERG pIC$_{50}$, BBB log([brain]:[blood]), BBB

category, HIA category, P-gp category, 2D6 affinity category, PPB$_{90}$ category and TPSA. Relevant ADMET properties were incorporated along with the docking predictions into multi-objective additive functions customized to the location of the target hepsin (circulatory system or urine). The values of the different properties were mapped to partial scores using affine desirability functions in StarDrop (Supporting Information Tables S1 and S2), and the different partial scores were summed to define the multi-objective functions "Blood Score" and "Urine Score".

3. Results

Hepsin, also referred to as TMPRSS1, was the first type II transmembrane serine protease to be cloned. It possesses an N-terminal cytosolic domain, a hydrophobic transmembrane region, and two extracellular domains bound together. The extracellular domains are cleaved after expression and are subsequently reconnected by a disulfide bond and a network of noncovalent interactions[37,38]. The larger, catalytic domain is structurally homologous to the trypsin family of serine proteases; whereas the smaller, non-catalytic domain (also called the stem region) adopts a scavenger-receptor cysteine-rich (SRCR) fold. The latter has poor sequence homology to other known sequences and its function is unclear, although it may serve to orient the catalytic protease domain[38]. The catalytic domain shows extensive sequence homology to other serine proteases like matriptase and HGFA.

Proteases contain several pockets (P1′, P1, P2, etc.) that participate in the recognition of specific amino acids in the substrate around the scissile bond (S1′, S1, S2, etc.) and confer them specificity[45,46]. Importantly, the primary specificity pocket S1 of hepsin contains an Asp189 and an Ala190 (Fig. S1); thus, the natural substrates for hepsin (and HGFA) contain a basic Arg in P1 position. It has also been reported that hepsin and HGFA have a preference for Leu in the P2 position[28]. By contrast, related matriptase and enteropeptidase have a Ser190 in the S1 pocket. Thus, the characteristics of both S1 and S2 pockets are key to maximize the specificity of inhibitors towards hepsin, the development of which becomes a particularly complex challenge.

The pipeline developed for this work is summarized in Fig. 2. The proprietary library used (4.84 × 10^6 conformers) included compounds available from a wide range of commercial vendors (ChemBridge, Asinex, Maybridge, Enamine, LifeChemicals, Specs, InterBioScreen, ChemDiv and KeyOrganics) and was virtually screened using the software LIDAEUS. 10,000 unique molecules were prioritized by rigid-body virtual screening on hepsin and then docked using Vina. This virtual screening may not be sensitive to fine details of binding associated to the binding pockets, which may require conformational changes of the ligand. Thus, the correlation between virtual screening and docking results downstream is weak (Supporting Information Fig. S2). However, all the ligands preselected by LIDAEUS would have binding constants in the submillimolar range according to Vina. Next, we docked the selected 10,000 compounds on matriptase and HGFA using Vina. The results are compiled in file mmc2.xlsx. The file also lists the predicted binding energies to matriptase and HGFA. By and large, the binding affinities to hepsin and matriptase show a higher correlation than the binding affinities to hepsin and HGFA, since the former belong to the same family of type II transmembrane serine proteases.

At this stage, we define the Specificity Index as Eq. (1):

$$\text{Specificity Index} = \frac{1/K_{\text{i, hepsin}}}{1/K_{\text{i, hepsin}} + 1/K_{\text{i, matriptase}} + 1/K_{\text{i, HGFA}}}$$
$$= \frac{e^{\Delta G \text{ hepsin}}}{e^{\Delta G \text{ hepsin}} + e^{\Delta G \text{ matriptase}} + e^{\Delta G \text{ HGFA}}}$$

(1)

Notice that $1/K_{\text{i}} = K_{\text{a}}$, the association constant. Thus, the Specificity Index quantitates how much ligand would bind, under equilibrium, to hepsin rather than to the related targets matriptase and HGFA, if all of them were equally accessible to the ligand and present at the same concentration.

Fig. 3 shows the Specificity Index as a function of the estimated free binding energy for the 10,000 compounds docked using Vina. The figure evidences that the compounds binding more strongly to hepsin are not necessarily the ones with the highest Specificity Index, which highlights the importance of explicitly considering potential off-target effects early on in the drug design process. Fortunately, a number of candidates show promise by having both high binding affinity and relatively high specificity towards hepsin.

Previous studies have shown that the combination of docking scores from different algorithms (consensus scoring) can significantly improve the accuracy of new drug designs. The combination of Vina and AutoDock, which use different force fields and scoring functions (see Methods), has proved particularly reliable[47−49]. Thus, we selected the top 1000 candidates based on their binding affinity to hepsin as predicted by Vina. Then, the binding of these compounds on hepsin, HGFA, and matriptase was independently re-evaluated using AutoDock. The results of these calculations are included in the file mmc2.xlsx. Vina and Auto-Dock show little agreement on the binding affinity of many ligands (Fig. 4a), to the point that the top-ranking compounds according to Vina would exhibit limited binding affinity according to AutoDock. We also see limited agreement between the Specificity Index values computed using binding energies from Vina and AutoDock (Fig. 4b), with AutoDock predicting somewhat more extreme specificities. We investigated whether this disagreement could be due to numerical artifacts, such as Auto-Dock getting trapped in local energy minima during the search and optimization algorithm. To this end, some molecules were repeatedly docked more than once starting from different conformations. Supporting Information Fig. S3 shows the coefficient of variation of the predicted binding energies, after three or more conformers of a ligand were docked on hepsin. The results evidence that both AutoDock and Vina converge to a final pose with the same binding energy within the termination tolerance of the search and optimization algorithm. This indicates that entrapment in local minima is not an issue in these results. Notably, AutoDock predicts that certain ligands may exhibit an even higher affinity than that predicted by Vina (Fig. 4b), and these deserve additional consideration (see below).

We evaluated the ADMET properties of the top 1000 candidates using StarDrop. The results are included in the file mmc2.xlsx. In order to summarize the ADMET characteristics of the ligands, we defined two multi-objective functions. The first function, Blood Score, evaluates the ADMET suitability of a drug meant to be administered orally and reach its target via the blood stream, which could be the case of targeting hepsin in the prostate. The components of this multi-objective function are indicated in Table S1 and were selected based on well-known oral lead- and
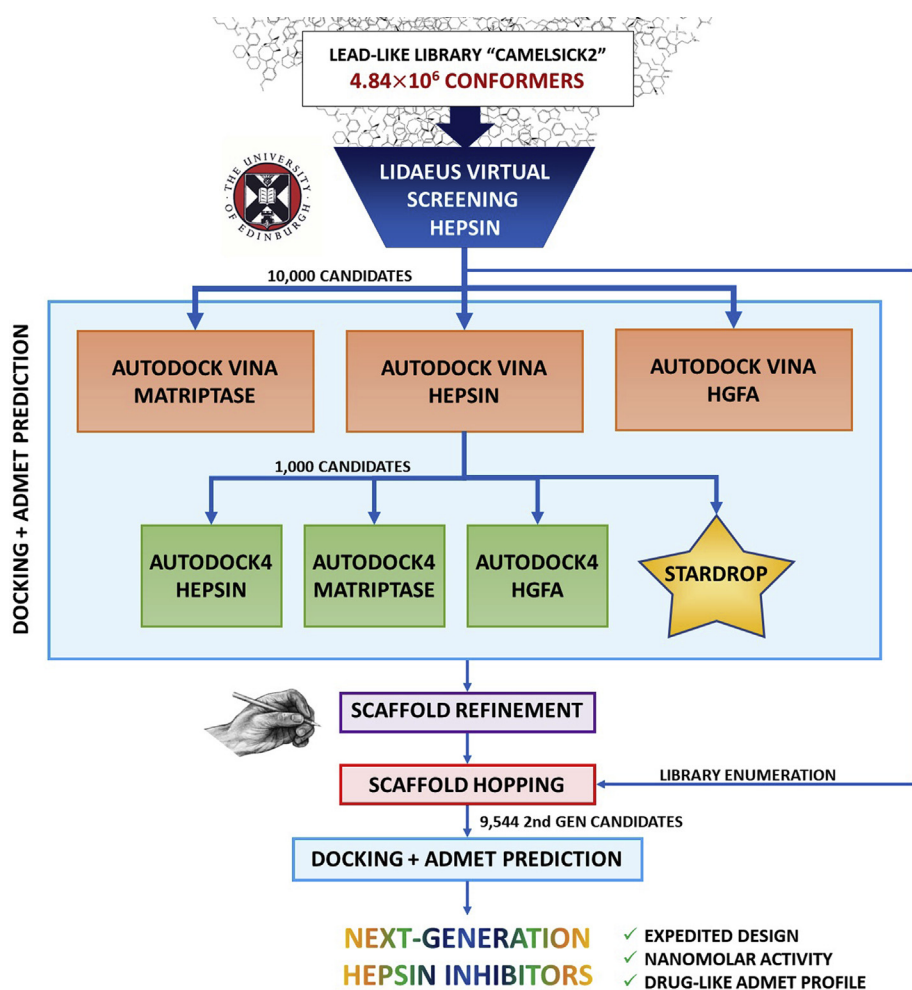
Figure 2 Hierarchical virtual screening methodology developed for this work.

drug-likeness criteria[50]. We modified this function in order to select for compounds with improved excretion and stability, aiming to inhibit hepsin exposed to urine (blood filtrate) in the nephron. To favor a somewhat faster excretion of the ligand, we increased the weights corresponding to solubility ($\log S$ and $\log P$). We also set a higher penalty for binding to undesired targets such as plasma protein, P-glycoprotein and hERG, crossing the brain−blood barrier, or having moderate affinity for the cytochromes P450 CYP2D6 and CYP2C9. This multi-objective function was called Urine Score, and its individual components are listed in Table S2. As illustrated in Supporting Information Fig. S4, the functions have a different behavior. Most molecules tend to rank higher in one of the two criteria and this might be exploited to target hepsin by different channels. Unfortunately, the ligands with the highest affinities for hepsin show moderate ADMET profiles, and thus it may be desirable to modify of some of their chemical groups and side chains.

The top-binding ligands according to AutoDock are listed in Table 1 (full details for any of the compounds can be looked up in file mmc2.xlsx). In this case, compound 31DRH1-100-176 is predicted to reach a very high $pK_i = 8.1$ ($\Delta G = -11.0$ kcal/mol) according to AutoDock, although this value was significantly lower in the predictions by Vina ($pK_i = 6.3$, $\Delta G = -8.6$ kcal/mol). The energy-minimizing poses for this ligand are presented in Fig. 5. Both programs situate the S-containing scaffold inside

the S1 pocket, but they differ in the orientation of the pendant chain. Notice also the presence of the terminal amide in position three of the fused ring system, which is able to establish strong hydrogen bonds with the backbone groups of Gly-216 and Gln-192.

In its turn, Supporting Information Table S3 shows the top five ligands in the original library based on their binding affinity to hepsin predicted by Vina as well as additional estimates of key properties. The top ligand based on its predicted binding affinity is 8DRH1-052-975, and it would show a pK_i of 7.0 ($\Delta G = -9.6$ kcal/mol). Its predicted specificity is relatively low and thus, *in vivo*, where it could be exposed to other targets, it might be outperformed by other ligands such as 6DRH1-317-399 or 31DRH1-219-368 in terms of its specific inhibition of hepsin. The binding pose of this ligand is shown in Supporting Information Fig. S5. Remarkably, the strongest interactions (hydrogen bonds) of this ligand are actually not being established inside of the S1 pocket, but outside of this pocket by the less constrained dihydroxyphenyl substituent. When the top ligand is docked on matriptase (Fig. 5), a very different pose is obtained for this molecule. In position 190 in matriptase, a serine residue is present, ready to interact as an HBD (Fig. S5). Therefore, the ligand would rotate when interacting with matriptase to bind even more strongly. Thus, in order to design a ligand specific for hepsin it becomes crucial to exploit this difference in position 190. Notice
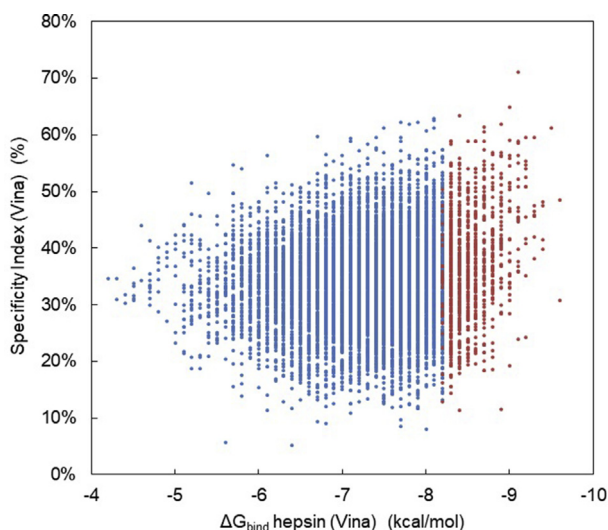
Figure 3 Relationship between specificity to hepsin and binding energy predicted by Vina for the compounds in the original library. The top 1000 hits based on their predicted binding energy to hepsin are highlighted in red.
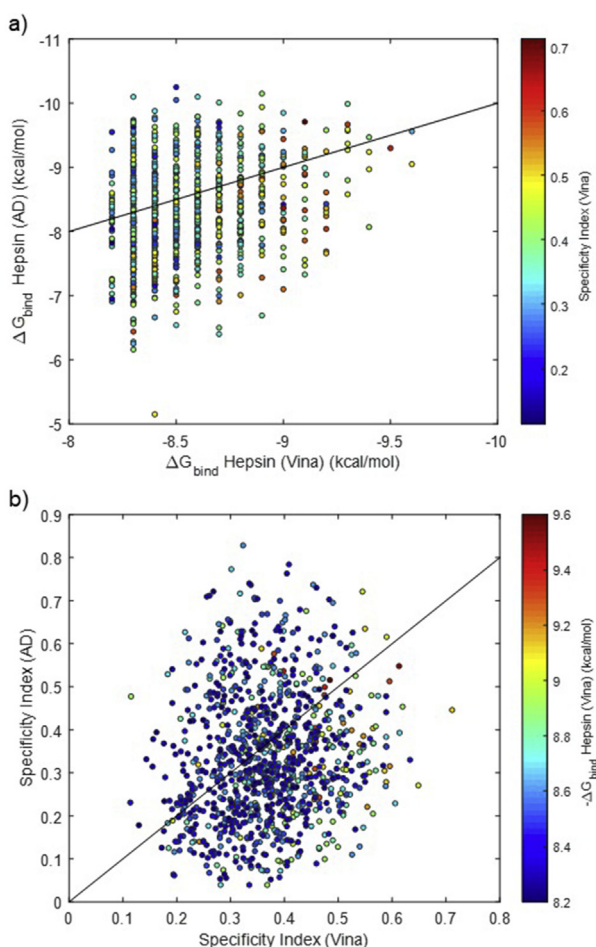

Figure 4 (a) Relationship between the Vina and AutoDock scores on hepsin for the top 1000 first-generation compounds. (b) Specificity indices were computed based on the results by Vina and AutoDock. The solid lines represent $y=x$.

that these important differences in binding would be invisible to traditional biochemical binding assays.

Notably, the top-binding compounds tend to present a linker bound by an amide, which can interact strongly with the residue Ser195 at the mouth of the pocket. Ser195 is, precisely, the key catalytic residue in the catalytic triad, and thus binding of these ligands would lead to the deactivation of the enzyme. An overlay of the top ligands on hepsin indicates that the side chain hydroxyl of Ser195 can be taken advantage by many ligands, either as HBD or as HBA (Supporting Information Fig. S6). The backbone amino group of Gly193 is also located such that it can interact with many linkers when the ligand's scaffold is in the S1 pocket.
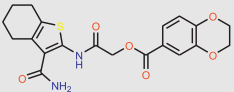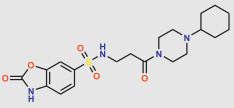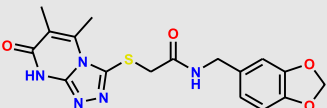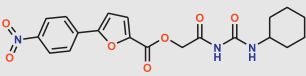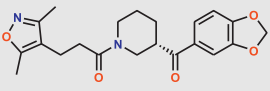
Overall, if we compare the top ligands predicted by AutoDock (Table 1) and Vina (Table S3) we observe some common scaffolds, such as 1,4-benzodioxan, 1-3-benzodioxole, and derivatives of purine. In general, it seems that systems of two fused five- and six-membered rings could be suitable scaffolds for designing potent hepsin inhibitors around its S1 pocket.

Based on the structural insights gained from the docking on hepsin, a small set of analogs of the top-binding ligands V-0 (Supporting Information Table S4) and VA-0 (Supporting Information Table S5) were manually designed. These molecules were docked using Vina and AutoDock, and their ADMET properties were predicted using StarDrop. Full results are included as Supporting Information file mmc3.xlsx. The virtual medicinal chemistry studies in Tables S4 and S5 highlight that i) is important to project side chains outside the S1 binding pocket to achieve submicromolar binding affinities and compensate for an excessive entropic loss upon binding, and ii) it is suitable to use a conjugated flat scaffold to design potent inhibitors, as these lead to a good spatial match with the shape of the pocket in hepsin and the orientation of the residues in the pocket.

Importantly, we identified that Asp189, which is crucial for the recognition of the native substrate, was not being fully exploited by the original top ligand candidates (Figs. 5 and S6). Furthermore, in the case of matriptase, Ser190 may act as HBA, whereas in hepsin Ala190 would not be able to engage in such interaction. In Table S4, we also explored the design of purine and caffeine-based scaffolds, with limited success, starting with a structural simplification approach[51]. For instance, we hypothesized that it could be beneficial to distance or substitute HBA groups posed to bind nearby position 190 in hepsin to improve the specificity of the ligand. Thus, we shifted the anchoring point of the side chain in 3-methylxanthine from 9-position (V-3) to 7-position (V-4), a change that could be synthetically accessible. To maximize the discrimination, we also considered introducing an amine group as an HBD (in V-12, V-13, V-15, and V-16) so that it clashed with Ser190 in matriptase. Despite these attempts, better results are reported in Table S5, where we started from the top ligand according to AutoDock (VA-0) and modified the indole-based scaffold to 5-fluoro-6-methyl-1H-indole (VA-14), boosting its affinity (from pK_i=6.2 to pK_i=7.2) and specificity (from 40% to 59%). Importantly, both AutoDock and Vina clearly agree that this molecule would achieve superior binding affinity and specificity for hepsin. Note that both purine and indole are examples of privileged scaffolds present in successful commercial drugs[50,52].

Another important detail of using the indole-based scaffold chosen is that it may nullify the potential proteolytic activity of the enzyme when the ligand contains an amide bond proximal to the 2-position of the indole system. The mechanism of action of serine proteases involves the nucleophilic attack of the carbonyl carbon

Table 1 Top five ligands in the original library based on their binding affinity to hepsin predicted by AutoDock and some of their properties.

| Structure	SPH identifier	pK_i/Specificity (Vina)	pK_i/Specificity (AD)	ADMET Blood Score/Urine Score	ADMET properties
	31DRH1-100-176 (VA-0)	6.3 42%	8.1 57%	0.101 0.008	MW = 416.4 Da LogS = 1.94 LogP = 2.24 2C9 pK_i = 5.23 hERG pIC_{50} = 3.86 P-gp category: yes PPB$_{90}$ category: medium
	29DRH1-560-653	6.2 21%	7.5 21%	0.271 0.093	MW = 436.5 Da LogS = 3.60 LogP = 1.86 2C9 pK_i = 4.17 hERG pIC_{50} = 5.47 P-gp category: yes PPB$_{90}$ category: low
	21DRH1-180-486	6.5 41%	7.4 48%	0.108 0.023	MW = 387.4 Da LogS = 2.20 LogP = 0.25 2C9 pK_i = 5.15 hERG pIC_{50} = 4.15 P-gp category: yes PPB$_{90}$ category: low
	31DRH1-041-844	6.3 35%	7.4 70%	0.082 0.012	MW = 415.4 Da LogS = 0.66 LogP = 3.62 2C9 pK_i = 4.71 hERG pIC_{50} = 4.05 P-gp category: yes PPB$_{90}$ category: high
	9DRH1-047-413	6.1 34%	7.4 45%	0.348 0.030	MW = 384.4 Da LogS = 1.89 LogP = 2.47 2C9 pK_i = 5.13 hERG pIC_{50} = 5.24 P-gp category: no PPB$_{90}$ category: medium
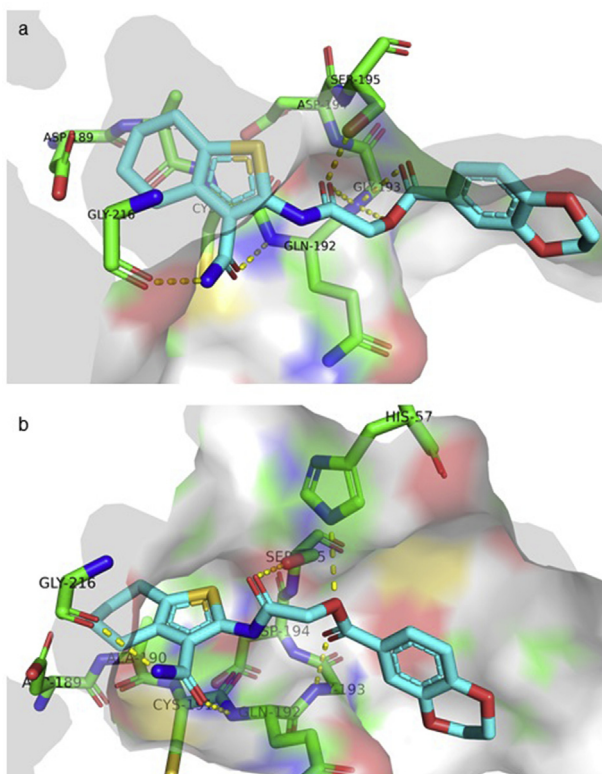
Figure 5 Binding poses of the ligand 31DRH1-100-176 (VA-0) on hepsin according to (a) AutoDock ($pK_i = 8.1$) and (b) Vina ($pK_i = 6.3$). Hydrogen-bonds are indicated as dashed yellow lines. A part of the pocket surface is shown: the exterior is lightly colored; the interior is rendered plain gray. Note the potential of Asp189 to act as HBA.

of the substrate peptide by a deprotonated serine[45,46] (Ser195). In this case, the proton on 1-nitrogen of indole is relatively acidic ($pK_a \sim 21$) and could be taken up by serine (Supporting Information Fig. S7), as it is also much closer and better oriented than the carbonyl carbon. This peculiar arrangement could further strengthen the binding of the ligand. Most importantly, because the directionality of the amide bond is inverted with respect to the natural substrate, the indole is sterically unable to act as a leaving group (as it is contained in the S1 pocket). This type of chemistry allows the powerful use of amide-containing inhibitors in close proximity to an otherwise catalytic serine as a notable example of substrate mimicry.

After optimizing the scaffold to 5-fluoro-6-methyl-1H-indole (compound VA-14, Table S5), we set out to enhance the side chain protruding outside the S1 pocket. A scaffold hopping approach was devised based on the original 10,000 ligands preselected by LIDAEUS. For this, we considered all the R-groups that were connected through an amide bond to the rest of the molecule in this first-generation library. Then, all the possible compounds conforming to Fig. 6 were enumerated, which resulted in a second-generation library of 9544 new hybrid molecules (file mmc4.xlsx). Thanks to the initial virtual screening by LIDAEUS, many of the side groups considered at this stage were known to fit reasonably well in the vicinity of the active site of hepsin, hence gaining a double benefit from the use of this software.

The new compounds were then docked using Vina and their ADMET properties estimated (file mmc4.xlsx). The most promising ligands were also docked on matriptase and HGFA in order

to assess their specificity. Although scaffold hopping onto a lead-like library might generate considerably large molecules (depending on how the R-group decomposition is conducted), most of the best molecules proposed satisfy Lipinski's rule of five[50] and exhibit reasonable ADMET properties. By and large, this library enumeration approach turned out very successful (Supporting Information Fig. S8) and allowed us to significantly upgrade the lead compound VA-14 ($K_i = 643$ nmol/l). Compounds with predicted pK_i values for hepsin as high as 8.1 ($K_i = 10$ nmol/L) were constructed by this approach (Table 2 and file mmc4.xlsx). The predicted binding poses of some of these inhibitors on hepsin are represented in Fig. 7 and Supporting Information Fig. S9 using LigPlot+ projections[53].

The binding affinities of the ligands in Table 2 were also estimated using a different scoring algorithm, X-score[54]. Although this scoring function presents some correlation with that in Vina, it is clearly different[42], and differences of two orders of magnitude in binding affinity to hepsin are predicted for other ligands (Supporting Information Fig. S10). Thus, the X-score results for the ligands selected, shown in Supporting Information Table S6, further suggest that these molecules would display high binding affinities to hepsin. Notably, some of the ligands designed in this work are predicted to be more potent than previous designs (Supporting Information Table S7), which could reduce the amounts of xenobiotics that need to be administered in order to achieve a therapeutic effect, potentially increasing safety and minimizing side-effects. A molecular pathway association analysis using PathwayMap[55] also suggests low toxicity for the compounds selected (Supporting Information Fig. S11). Notably, the compounds selected have molecular weights in the range of 400 Da (well below Lipinski's 500 Da cutoff), ≤ 10 rotatable bonds, and a polar surface area ≤ 140 Å^2 (Table 2). According to Veber's rule, inhibitors that meet the latter two criteria show good oral bioavailability[56].

An additional advantage of having started with a lead-like filtered library is that most of the compounds show good synthetic accessibility[57] and many of them are commercially available. Likewise, because the second-generation library was enumerated by recombining R-groups at amide bonds, the molecules built generally have reasonable synthetic accessibility. Notably, the optimized scaffold 5-fluoro-6-methyl-1H-indole can be acquired from commercial vendors. In Supporting Information Fig. S12 we
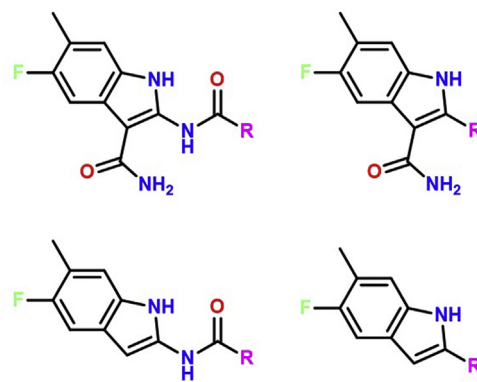

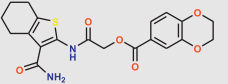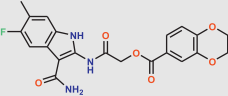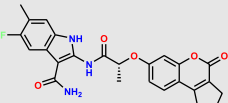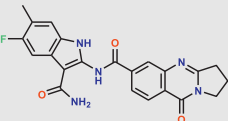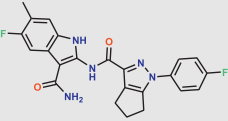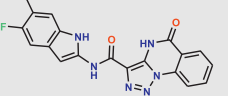**Figure 6** 9544 new hybrid compounds (second-generation library) were generated by extracting R-groups from the LIDAEUS virtual screening output (original library) and exhaustively combining them with the optimized scaffold in the four ways shown. For details, see file mmc4.xlsx.

Table 2 Summary of the development of specific drug-like hepsin inhibitors in this work.

| Library	Structure	Identifier	K_i (nmol/L)/S.I. (Vina)	K_i (nmol/L)/S.I. (AD)	Blood Score/Urine Score/Rot. bonds/PSA	ADMET properties
1st Gen library		8DRH1-052-975 (VA-0)	578 40%	12 53%	0.101 0.008 7 117	MW = 416.4 $\log S$ = 1.9 $\log P$ = 2.24 2C9 pK_i = 5.23 hERG pIC_{50} = 3.86
Manual refinement		VA-14	64 59%	31 55%	0.061 0.005 7 133	MW = 427.4 $\log S$ = 0.87 $\log P$ = 2.15 2C9 pK_i = 5.70 hERG pIC_{50} = 3.49
2nd Gen library		VB-1013	10 75%	66 14%	0.028 0.005 6 127	MW = 463.5 $\log S$ = 0.63 $\log P$ = 3.99 2C9 pK_i = 6.04 hERG pIC_{50} = 4.13
		VB-1385	10 64%	13 52%	0.091 0.008 4 123	MW = 419.4 $\log S$ = 0.91 $\log P$ = 2.79 2C9 pK_i = 5.73 hERG pIC_{50} = 4.23
		VB-3497	10 60%	37 38%	0.107 0.022 5 106	MW = 435.4 $\log S$ = 0.49 $\log P$ = 3.65 2C9 pK_i = 5.99 hERG pIC_{50} = 4.52
		VB-865	17 55%	18 58%	0.159 0.022 3 108	MW = 376.3 $\log S$ = 0.52 $\log P$ = 2.57 2C9 pK_i = 5.81 hERG pIC_{50} = 4.77
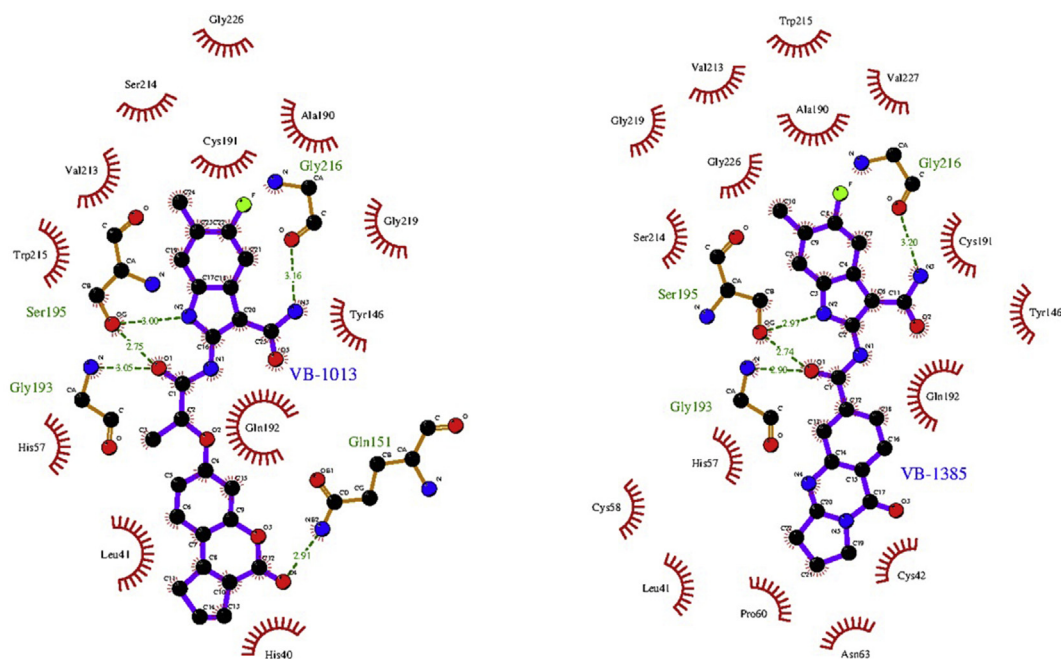
Figure 7 Projections of the inhibitors VB-1013 (left) and VB-1385 (right) bound to the active site of hepsin as predicted by Vina. The pipeline proposed succeeded in incorporating side chains to the scaffold that make strong interactions with hepsin and are synthetically accessible.

exemplify the synthesis of the most promising ligands in Table 2 starting from a commercially available compound.

4. Conclusions

Hepsin is an important medicinal target in need of improved inhibitors. The development of drug-like specific inhibitors of hepsin could improve our understanding and options about this macromolecule in kidney stone formation and the progression of certain types of cancer. Known inhibitors are limited by their specificity towards hepsin in the presence of similar proteases and/or by their limited stability in the human body. In this work, potentially improved small-molecule inhibitors have been designed to target hepsin in a specific manner. Derivatives of indole, a privileged scaffold, were predicted to exhibit nanomolar binding affinity to the active site of hepsin. Moreover, their estimated ADMET properties are promising towards oral administration.

At a broader level, this work defines a hierarchical computational pipeline that can allow the determination of potent lead-like and drug-like compounds with excellent cost-efficiency. The combination of a coarser virtual screening prior to a finer docking and the recombination of the results in a second-generation library allowed three important benefits: i) a large lead-like library could be reduced into a compact library of compounds with sizes matching the environment of the active site of the protein of interest, ii) the second-generation library takes further advantage of the first virtual screening, and iii) the compounds produced are more likely to exhibit reasonable ADMET properties and better synthetic accessibility. Moreover, the pipeline proposed can be applied with little intervention by advancing the top hits obtained by default in the pipeline, or it can be adapted to the know-how of the laboratory or company using it, allowing the manual modification of intermediate structures, as shown in this work. Notably, over a million compounds were screened to varying depths

without the need of expensive supercomputing resources, by using an affordable 32-core system in less than a week of computing time. This work could therefore be adapted to pursue novel targets and improve the efficiency of the discovery and design of next-generation drugs.

Author contributions

Vincent Blay designed the research. Vincent Blay and Douglas R. Houston conducted the simulations. Vincent Blay, Douglas R. Houston and Mu-Chun Li analyzed the results and wrote the paper. Hsing-Pang Hsieh, Sunita P. Ho, and Mashall L. Stoller reviewed the manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Appendix A. Supporting information

Supporting data to this article can be found online at https://doi.org/10.1016/j.apsb.2019.09.008.

References

1. Ziemba JB, Matlaga BR. Epidemiology and economics of nephrolithiasis. *Investig Clin Urol* 2017;**58**:299−306.

2. Sohgaura A, Bigoniya P. A review on epidemiology and etiology of renal stone. *Am J Drug Discov Dev* 2017;**7**:54−62.

3. Alelign T, Petros B. Kidney stone disease: an update on current concepts. *Adv Urol* 2018;**2018**:3068365.

4. Grases F, Costa-Bauzá A. Mechanisms of renal and salivary. Calculi formation and development. In: Königsberger E, Königsberger LC, editors. *Biomineralization−medical aspects of solubility*. New York, NY: Wiley; 2006. p. 39−123.

5. Dussol B, Berland Y. Urinary kidney stone inhibitors. Where are we?. *Nephrol Dial Transplant* 1996;**11**:1222−4.

6. Basavaraj DR, Biyani CS, Browning AJ, Cartledge JJ. The role of urinary kidney stone inhibitors and promoters in the pathogenesis of calcium containing cenal stones. *EAU-EBU Update Ser* 2007;**5**:126−36.

7. Aggarwal KP, Narula S, Kakkar M, Tandon C. Nephrolithiasis: molecular mechanism of renal stone formation and the critical role played by modulators. *BioMed Res Int* 2013:292953. 2013.

8. Monico CG, Milliner DS. Genetic determinants of urolithiasis. *Nat Rev Nephrol* 2012;**8**:151−62.

9. Edvardsson VO, Goldfarb DS, Lieske JC, Beara-Lasic L, Anglani F, Milliner DS, et al. Hereditary causes of kidney stones and chronic kidney disease. *Pediatr Nephrol* 2013;**28**:1923−42.

10. Vasudevan V, Samson P, Smith AD, Okeke Z. The genetic framework for development of nephrolithiasis. *Asian J Urol* 2017;**4**:18−26.

11. Rashed T, Menon M, Thamilselvan S. Molecular mechanism of oxalate-induced free radical production and glutathione redox imbalance in renal epithelial cells: effect of antioxidants. *Am J Nephrol* 2004;**24**:557−68.

12. Wang L, Nancollas GH. Dynamics of biomineralization and biodemineralization. In: Sigel A, Sigel H, Sigel RKO, editors. *Biomineralization: from nature to application*, vol. 4. New York, NY: Wiley; 2008. p. 413−56.

13. Khan SR, Pearle MS, Robertson WG, Gambaro G, Canales BK, Doizi S, Traxer O, Tiselius HG. Kidney stones. *Nat Rev Dis Primers* 2016;**2**:16008.

14. Matlaga BR, Coe FL, Evan AP, Lingeman JE. The role of Randall's plaques in the pathogenesis of calcium stones. *J Urol* 2007;**177**:31−8.

15. Khan SR, Canales BK. Unified theory on the pathogenesis of Randall's plaques and plugs. *Urolithiasis* 2015;**43**:S109−23.

16. Santambrogio S, Cattaneo A, Bernascone I, Schwend T, Jovine L, Bachi A, et al. Urinary uromodulin carries an intact ZP domain generated by a conserved C-terminal proteolytic cleavage. *Biochem Biophys Res Commun* 2008;**370**:410−3.

17. Bokhove M, Nishimura K, Brunati M, Han L, de Sanctis D, Rampoldi L, et al. A structured interdomain linker directs self-polymerization of human uromodulin. *Proc Natl Acad Sci USA* 2016;**113**:1552−7.

18. Mo L, Huang HY, Zhu XH, Shapiro E, Hasty DL, Wu XR. Tamm-Horsfall protein is a critical renal defense factor protecting against calcium oxalate crystal formation. *Kidney Int* 2004;**66**:1159−66.

19. Devuyst O, Olinger E, Rampoldi L. Uromodulin: from physiology to rare and complex kidney disorders. *Nat Rev Nephrol* 2017;**13**:525−44.

20. Serafini-Cessi F, Malagolini N, Cavallone D. Tamm-Horsfall glycoprotein: biology and clinical relevance. *Am J Kidney Dis* 2003;**42**:658−76.

21. Mo L, Zhu XH, Huang HY, Shapiro E, Hasty DL, Wu XR. Ablation of the Tamm-Horsfall protein gene increases susceptibility of mice to bladder colonization by type 1-fimbriated *Escherichia coli*. *Am J Physiol Renal Physiol* 2004;**286**:F795−802.

22. Tokonami N, Takata T, Beyeler J, Ehrbar I, Yoshifuji A, Christensen EI, et al. Uromodulin is expressed in the distal convoluted tubule, where it is critical for regulation of the sodium chloride cotransporter NCC. *Kidney Int* 2018;**94**:701−15.

23. Viswanathan P, Rimer JD, Kolback AM, Ward MD, Kleinman JG, Wesson JA. Calcium oxalate monohydrate aggregation induced by aggregation of desialylated Tamm-Horsfall protein. *Urol Res* 2011;**39**:269−82.

24. Rambausek M, Dulawa J, Jann K, Ritz E. Tamm Horsfall glycoprotein in diabetes mellitus: abnormal chemical composition and colloid stability. *Eur J Clin Invest* 1988;**18**:237−42.

25. Hess B, Zipperle L, Jaeger P. Citrate and calcium effects on Tamm-Horsfall glycoprotein as a modifier of calcium oxalate crystal aggregation. *Am J Physiol* 1993;**265**:F784−91.

26. Brunati M, Perucca S, Han L, Cattaneo A, Consolato F, Andolfo A, et al. The serine protease hepsin mediates urinary secretion and polymerisation of Zona Pellucida domain protein uromodulin. *Elife* 2015;**4**:e08887.

27. Coleman WB. Molecular pathogenesis of prostate cancer. In: Coleman WB, Tsongalis GJ, editors. *Molecular pathology*. 2nd ed. Academic Press; 2018. p. 555−68.

28. Damalanka VC, Han Z, Karmakar P, O'Donoghue AJ, La Greca F, Kim T, et al. Discovery of selective matriptase and hepsin serine protease inhibitors: useful chemical tools for cancer cell biology. *J Med Chem* 2019;**62**:480−90.

29. Tanimoto H, Yan Y, Clarke J, Korourian S, Shigemasa K, Parmley TH, et al. Hepsin, a cell surface serine protease identified in hepatoma cells, is overexpressed in ovarian cancer. *Cancer Res* 1997;**57**:2884−7.

30. Sloane BF, List K, Fingleton B, Matrisian L. Proteases in cancer: significance for invasion and metastasis. In: Brix K, Stöcker W, editors. *Proteases: structure and function*. Vienna: Springer; 2013. p. 491−550.

31. Klezovitch O, Chevillet J, Mirosevich J, Roberts RL, Matusik RJ, Vasioukhin V. Hepsin promotes prostate cancer progression and metastasis. *Cancer Cell* 2004;**6**:185−95.

32. Wu Q, Yu D, Post J, Halks-Miller M, Sadler JE, Morser J. Generation and characterization of mice deficient in hepsin, a hepatic transmembrane serine protease. *J Clin Investig* 1998;**101**:321−6.

33. Yu IS, Chen HJ, Lee YS, Huang PH, Lin SR, Tsai TW, et al. Mice deficient in hepsin, a serine protease, exhibit normal embryogenesis and unchanged hepatocyte regeneration ability. *Thromb Haemost* 2000;**84**:865−70.

34. Kwon H, Han J, Lee KY, Son SH, Byun Y. Recent Advances of hepsin-targeted inhibitors. *Curr Med Chem* 2017;**24**:2294−311.

35. Damalanka VC, Janetka JW. Recent progress on inhibitors of the type II transmembrane serine proteases, hepsin, matriptase and matriptase-2. *Future Med Chem* 2019;**11**:743−69.

36. Torres-Rosado A, O'Shea KS, Tsuji A, Chou SH, Kurachi K. Hepsin, a putative cell-surface serine protease, is required for mammalian cell growth. *Proc Natl Acad Sci U S A* 1993;**90**:7181−5.

37. Hooper JD, Clements JA, Quigley JP, Antalis TM. Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes. *J Biol Chem* 2001;**276**:857−60.

38. Somoza JR, Ho JD, Luong C, Ghate M, Sprengeler PA, Mortara K, et al. The structure of the extracellular region of human hepsin reveals a serine protease domain and a novel scavenger receptor cysteine-rich (SRCR) domain. *Structure* 2003;**11**:1123−31.

39. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J* 2008;**275**:1−21.

40. Taylor P, Blackburn E, Sheng YG, Harding S, Hsin KY, Kan D, et al. Ligand discovery and virtual screening using the program LIDAEUS. *Br J Pharmacol* 2008;**153**:S55−67.

41. Houston DR, Yen LH, Pettit S, Walkinshaw MD. Structure- and ligand-based virtual screening identifies new scaffolds for inhibitors of the oncoprotein MDM2. *PLoS One* 2015;**10**:e0121424.

42. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem* 2010;**31**:455−61.

43. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 1998;**19**:1639−62.

44. Jaghoori MM, Bleijlevens B, Olabarriaga SD. 1001 Ways to run AutoDock Vina for virtual screening. *J Comput Aided Mol Des* 2016;**30**:237−49.

45. Hedstrom L. Serine protease mechanism and specificity. *Chem Rev* 2002;**102**:4501−23.

46. Blay V, Pei D. Serine proteases: how did chemists tease out their catalytic mechanism?. *ChemTexts* 2019. Available from: https://doi.org/10.1007/s40828-019-0093-4.

47. Kukol A. Consensus virtual screening approaches to predict protein ligands. *Eur J Med Chem* 2011;**46**:4661−4.

48. Chang MW, Ayeni C, Breuer S, Torbett BE. Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. *PLoS One* 2010;**5**:e11955.

49. Houston DR, Walkinshaw MD. Consensus docking: improving the reliability of docking in a virtual screening context. *J Chem Inf Model* 2013;**53**:384−90.

50. Blass B, editor. *Basic principles of drug discovery and development*. Academic Press; 2015.

51. Wang S, Dong G, Sheng C. Structural simplification: an efficient strategy in lead optimization. *Acta Pharm Sin B* 2019;**9**:880−901.

52. Welsch ME, Snyder SA, Stockwell BR. Privileged scaffolds for library design and drug discovery. *Curr Opin Chem Biol* 2010;**14**:347−61.

53. Laskowski RA, Swindells MB. LigPlot+: multiple ligand−protein interaction diagrams for drug discovery. *J Chem Inf Model* 2011;**51**:2778−86.

54. Wang R, Lai L, Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* 2002;**16**:11−26.

55. Jiménez J, Sabbadin D, Cuzzolin A, Martínez-Rosell G, Gora J, Manchester J, et al. PathwayMap: molecular pathway association with self-normalizing neural networks. *J Chem Inf Model* 2019;**59**:1172−81.

56. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 2002;**45**:2615−23.

57. Lin FY, Esposito EX, Tseng YJ. LeadOp+R: structure-based lead optimization with synthetic accessibility. *Front Pharmacol* 2018;**9**:96.