



Furin inhibitors: Importance of the positive formal charge and beyond

Fabian López-Vallejo, Karina Martínez-Mayorga *

Torrey Pines Institute for Molecular Studies, 11350 SW Village Parkway, Port St. Lucie, FL 34987, USA

ARTICLE INFO

Article history:

Received 10 March 2012

Revised 3 May 2012

Accepted 12 May 2012

Available online 19 May 2012

Keywords:

Proprotein convertases

ChEMBL database

Molecular docking

ABSTRACT

Furin is the prototype member of the proprotein convertases superfamily. Proprotein convertases are associated with hormonal response, neural degeneration, viral and bacterial activation, and cancer. Several studies over the last decade have examined small molecules, natural products, peptides and peptide derivatives as furin inhibitors. Currently, subnanomolar inhibition of furin is possible. Herein, we report the analysis of 115 furin inhibitors reported in the literature. Analysis of the physicochemical properties of these compounds highlights the dependence of the inhibitory potency with the total formal charge and also shows how the most potent (peptide-based) furin inhibitors have physicochemical properties similar to drugs. In addition, we report docking studies of 26 furin inhibitors using Glide XP. Inspection of binding interactions shows that the two putative binding modes derived from our study are reasonable. Analysis of the binding modes and protein–ligand interaction fingerprints, used here as post-docking procedure, shows that electrostatic interactions predominate on S1, S2 and S4 subsites but are seldom in S3. Our models also show that the benzimidamide group, present in the most active inhibitors, can be accommodated in the S1 subsite. These results are valuable for the design of new furin inhibitors.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Proprotein convertases (PC) are in the class of calcium-dependent serine endoproteases responsible for converting inactive protein precursors into their active forms.^{1,2} PCs are involved in critical biological processes, including embryo formation,³ blood sugar and insulin levels, and hormonal control mechanisms. Pathological conditions linked to PC's misregulation include: obesity, diabetes, Alzheimer's disease, indications of tumor metastasis, etc.^{4–6} Thus, inhibition of PCs may be of critical therapeutic importance.

Furin, the prototype PC, has been recently crystallized, and resolved at 2.6 Å resolution.⁷ The overall structure of furin is shown in Figure 1A. Calcium ions revealed in the crystallographic structure are shown as orange spheres. Important features of the furin crystal structure include: (1) the key residues Ser368, His194, and Asp153, forming the active site triad arranged in the center; (2) a 'canyon-like' crevice, housing the triad; and (3) the inhibitor Dec-RVKR-cmk, covalently bound to furin. Nearby the triad, the catalytic site is defined by subsite positions, named S1, S2, etc., which are the counterparts of segments of the co-crystallized inhibitor, P1, P2, etc., following the S and P nomenclature proposed by Schechter and Berger.⁸ The molecular surfaces of the S subsites are depicted in Figure 1B. Note that subsites are typically composed for more than one residue. The interactions observed in the X-ray crystal structure of furin with the co-crystallized ligand are informative

for molecular modeling studies. Moreover, based on the furin crystal structure, homology models of other prohormone convertases have provided information on important similarities and differences between their binding sites⁹ and have also been used to hypothesize the location of allosteric binding sites.¹⁰

Strategies to inhibit PCs have been recently reviewed.¹¹ Potent furin inhibitors have been identified from natural products as well as from the structural modification of molecular scaffolds.^{6,12–16} Nanomolar inhibitors have been derived from, or inspired in combinatorial libraries.^{12,14,17} The large amount of peptides that can be accessible by means of combinatorial chemistry allows the generation of more complex and chiral compounds than those represented by small molecules.¹⁸ In particular, polyarginines (obtained from combinatorial libraries) and peptide derivatives have been increasingly studied in the last decade as potent inhibitors of furin.^{12,14,17,19} Initially, reported inhibition values were in the low-micromolar to mid-nanomolar ranges; recently, improved structures have yielded to low-nanomolar to sub-nanomolar potency.^{12,14,19} Therefore, peptides remain attractive starting points for the discovery of hits in drug discovery campaigns. In addition, the bioavailability, stability and selectivity of peptides can be improved by the formation of cyclic peptide structures. In fact, cyclic peptides from natural sources have shown to be promising on a variety of biological systems and the design^{20,21} and evaluation²² of cyclic peptide derivatives is an area of intense research. As part of our efforts to identify biologically-relevant ligands, we have explored the integration of virtual and experimental screenings.²³ In particular, we observed differentiations in docking scores among combinatorial libraries towards furin.²⁴ Greater understanding of

* Corresponding author.

E-mail address: kmartinez@tpims.org (K. Martínez-Mayorga).

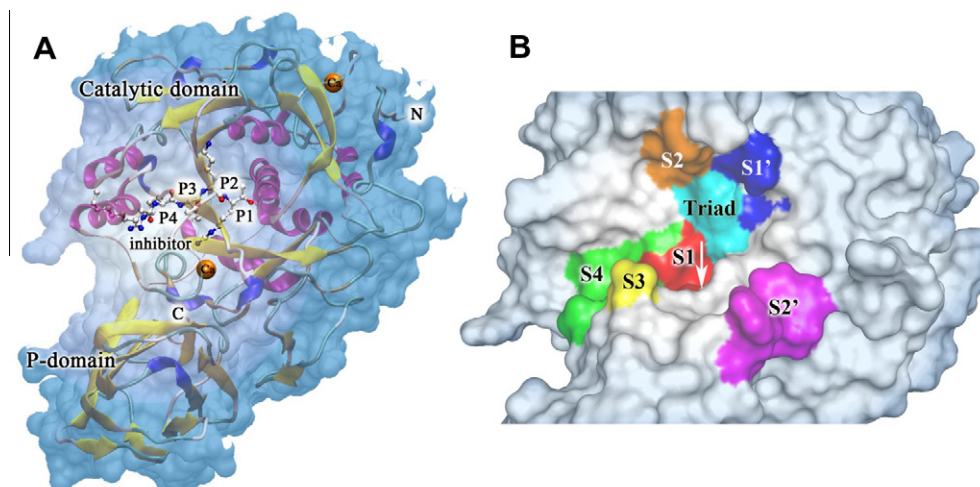


Figure 1. Molecular representation of furin. (A) Overall view: The catalytic domain is at the top with alpha-helical and beta-sheets and the P-domain are at the bottom with beta-sheets. Calcium ions are shown as orange spheres. The co-crystallized ligand is shown in ball and stick representation and colored by atom type and labeled by P group. (B) Residues forming the triad surface are shown cyan (Asp153, His194, and Ser368); residues forming the surface of subsites S1–S4 are shown in red, orange, yellow and green, respectively. Subsites S1' and S2' are also shown, colored in blue and magenta, respectively.

the pharmacophoric features of furin inhibitors is valuable for the design of new ligands as well as targeted combinatorial libraries.

The selection and understanding of computational methodologies are key aspects for the proper production, analysis and exploration of chemical compounds with relevance in Medicinal Chemistry. These aspects are described in the literature. For example, recent reports summarize the scope and pitfalls in virtual screening²⁵ and QSAR studies.²⁶ One of the commonly used methodologies in Medicinal Chemistry studies involving computational methods is molecular docking. Authoritative reviews on docking methods are reported elsewhere.^{27,28} Particularly, the wishful thinking of obtaining a linear correlation between docking energies and biological activity values is frequently manifested and has been revised.^{29,30}

Molecular models and molecular representations can be produced at different levels of sophistication. Examples in the chemoinformatics arena are: fingerprint representations, routinely used for similarity searches,³¹ and the visualization of the chemical space³² for the quick comparison of large compound collections. Reviews describing the many aspects, methodologies, and applicability of chemoinformatic analysis can be found in the literature.^{33–35}

In this study, we conducted a chemoinformatic analysis of compounds with known biological activity towards furin. Based on this analysis, molecules were selected for molecular docking studies.

2. Results and discussion

A total of 115 structures was collected from the ChEMBL database and from the literature,^{12,19,36–40} and are provided as [Supplementary data](#). This set contains 72 peptides and peptide derivatives and 43 small molecules. We will refer to the set comprising 72 peptides and peptide derivatives as peptides, unless otherwise noted. The 43 compounds referred here as small molecules have a non-peptidic nature. In addition, only those compounds with accurate evaluation of the biological activity were selected. The statistics of the experimental inhibition constant, K_i , values are summarized in [Table 1](#). The majority of the peptides have K_i values around 65.5 nM (median), with a mean more than 10 times higher (707 nM). This is also reflected in the high positive skewness, suggesting that there are only a few peptides with very large K_i values. The large skewness of this distribution is accompanied by a very large positive kurtosis. Kurtosis above three defines a double

exponential curve, signifying lower probability than a normally distributed variable for extreme values. These results indicate that, in this set, the proportion of extremely weak furin peptide inhibitors is very low, with respect to the median. The length of the peptides evaluated ranged from 1 to 9 amino acids. A frequency count of the amino acids in the peptide sequences provided the following information: (1) the more abundantly explored amino acids are: arginine, glycine, and alanine, with a maximum frequency count per peptide of 9, 6, and 4, respectively, (2) threonine, lysine, valine, and leucine were employed a maximum of two times in the same peptide, and the rest of the amino acids were employed a maximum of once per peptide, and (3) remarkably, arginine was present on average 3.8 times per peptide followed by lysine, employed on average 1.2 times per peptide. In the case of the small molecules, the difference between mean and median K_i values is even larger (more than 30 times), shown in [Table 1](#). Skewness and kurtosis follow the same trend than those for the furin peptide inhibitors. Noteworthy, the mean and median values (48073 and 1550 nM, respectively) are much higher than those for the peptide inhibitors. Thus, the majority of the small molecules evaluated towards furin have K_i values close to 1550 nM, and on average the small molecules evaluated were less active towards furin than the peptides. These rather simple statistical descriptors characterize the distributions shown underneath [Table 1](#) and can be taken into account when designing new furin inhibitors.

It should be pointed out that the value of this analysis is more apparent when analyzing a number of large distributions. For example, for collections classified by type of functional group, scaffold, etc., the descriptive statistics may provide a rough idea of the frequency and potency that such type of molecules have on a particular assay. This, of course, is not limited to binding affinity, as it also applies to other experimental measures; for instance, weighted values of selectivity, agonism, or antagonism.

2.1. Physicochemical properties of compounds evaluated against furin

Different molecular properties have been identified to characterize drug-likeness,⁴¹ lead likeness,⁴² launched compounds,⁴³ etc. In this work, we chose six drug-like physicochemical properties: molecular weight (MW), number of rotatable bonds (RB), the octanol/water partition coefficient ($SlogP$), topological polar

Table 1
Statistics of reported furin inhibitors

	Peptides	Small molecules
# Molecules	72	43
Range of K_i values (nM)	0.81–9710	6–500,000
Skewness	3.7	3.0
Kurtosis	14.3	9.7
Mean	707	48073
Median	65.5	1550

Peptides

Small molecules

surface area (TPSA), hydrogen bond acceptors (HBA), and hydrogen bond donors (HBD).

A comparison of small molecules and peptides based on these physicochemical properties is shown in Figure 2; as a reference, marketed drugs from DrugBank are also included. This property space was carried out by principal component analysis (PCA). Drugs are shown as gray squares, peptides as circles and small molecules as triangles. Data points are color-coded by biological activity in a binary manner representing active (green) and inactive (red) compounds. The threshold used for this classification,

100 nM, has been used before in a large experimental study.³⁷ Thus, compounds with K_i values, equal or lower than 100 nM ($pK_i = 7.0$) were considered actives, and compounds with K_i above 100 nM were considered as inactive. Based on this partition, 46 out of 72 peptides and 7 out of 43 small molecules were considered actives. In addition, data points are size-coded, the larger the shape the higher the content of positive charges.

Not surprisingly, the property space of the small molecules, located in sparse areas, is within or nearby that of drugs, (Fig. 2). A noticeable trend in the property space, shown as circles at the top center of Figure 2, corresponds to six polyarginines. This expected trend is a consequence of the homologue nature of these peptides. The first circle (far left) corresponds to the tetra-arginine while the circle to the far right corresponds to the nona-arginine. Polyarginines with three, four, and six arginine residues (six positive formal charges) fall in the inactive category, according to our threshold. Peptides with seven, eight, and nine arginine residues were actives. Moreover, this trend is observed not only when the data is analyzed in a categorical manner (active/inactive), but also when looking at the actual activity values.¹² Clearly, as the basicity increases (residues or R groups positively charged) the probability of finding active molecules also increases.¹²

Statistics for the six scaled molecular properties considered in this analysis are summarized in Table 2. On average, physicochemical properties of peptide inhibitors differ to those for small molecules inhibitors and drugs. For instance, the average MW of peptides (1200) is more than three times the MW of drugs and small molecules, 443 and 315, respectively. Average values for all the other properties are also higher for peptides. Further differentiation in physicochemical properties was observed based on active versus inactive peptides. On average, active peptides resemble the physicochemical properties of the group of peptides located far away from drugs in Figure 2. This is apparent by looking at the ratio of active versus inactive peptides close and far away from drugs in Figure 2. The statistics of peptides near to drugs, classified as actives and inactive, are also shown in Table 2. This leads to two interrelated observations: first, it is more challenging to find

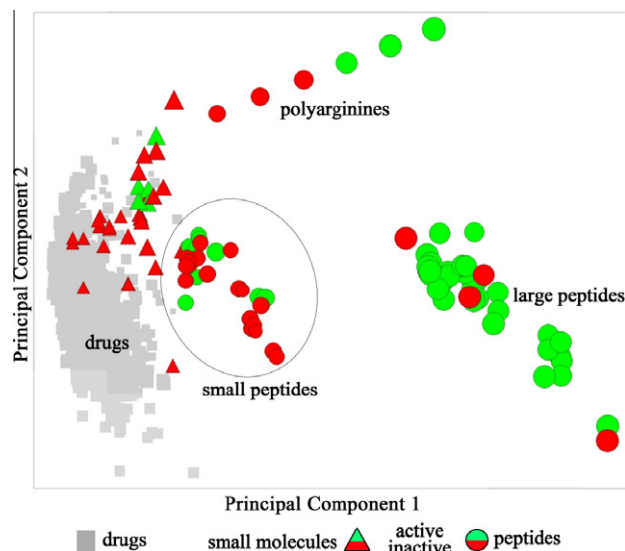


Figure 2. Property space of peptides and small molecules inhibitors of furin. Property space of drugs (DrugBank) is shown in gray squares, peptides are represented by circles, and small molecules by triangles. Active compounds are in green and inactive in red. Data points are size-coded by formal charge. The first two principal components accounts for 90.01% of the variance.

Table 2
Physicochemical properties distribution of peptides and small molecules inhibitors of furin

		DrugBank	Small molecules	Peptides	Peptides		
						All	Near drugs
HBD	Median	2	1	7.5	Active	12	4
					Inactive	5	5
	Mean	2.02	0.79	9.1	Active	11	4
					Inactive	6	5
	SD	1.81	1.08	4.77	Active	5	1
HBA	Median	3	2	8.5	Inactive	4	1
					Active	13	4
	Mean	3.58	2.65	9.71	Inactive	5	5
					Active	12	5
	SD	1.92	1.31	4.99	Inactive	7	6
SlogP	Median	2.47	−5.36	−20	Active	5	1
					Inactive	4	2
	Mean	2.37	−5.27	−15.62	Active	−22	−7
					Inactive	−7	−6
	SD	2.2	2.68	8.59	Active	−19	−6
TPSA	Median	68.53	255.14	732.1	Inactive	−10	−6
					Active	8	2
	Mean	72.41	236.82	601.43	Inactive	7	1
					Active	787	307
	SD	41.22	82.49	255.81	Inactive	371	336
RB	Median	5	12	61.5	Active	705	324
					Inactive	429	333
	Mean	5.05	10.86	54.11	Active	230	47
					Inactive	199	53
	SD	3.29	4.11	20.73	Active	66	31
MW	Median	313.79	449.57	1304	Inactive	37	35
					Active	63	33
	Mean	315.15	443.54	1200	Inactive	40	34
					Active	19	8
	SD	93.53	146.23	441.3	Inactive	16	7
					Active	1460	690
					Inactive	804	780
					Active	1379	758
					Inactive	901	771
					Active	397	131
					Inactive	341	121

active peptides in the region close to drugs, and second, in this region of the property space; these physicochemical properties cannot be used as a criterion to filter out inactive peptides.

Since furin subsites are all negatively charged, compounds with properly distributed basic residues (like arginine and lysine) or R groups positively charged have more complementarity with the binding site. Figure 3 shows the count of positive formal charges for peptides (A), and small molecules (B), and are differentiated as active (green) and inactive (red) ligands. A clear dependence of the activity with the positive formal charge was observed, coinciding with previous observations.¹² The dependency of the activity with the formal charge for both sets combined is shown on Figure 3C. The predominance of active compounds with formal charge equal or greater than four is notorious. Interestingly, these active furin inhibitors did not reach the highest inhibition found across the set ($pK_i > 8$). Furthermore, five out of eight of the most active compounds have less than four positive formal charges. This fact demonstrates how, besides incrementing the formal charge, refinement of other factors, like the spatial complementarity in the binding site, can yield highly potent furin inhibitors.

The dependency of the binding affinity to furin with the number of positive charges was observed before¹² on a smaller set of furin inhibitors. This observation becomes apparent by examining furin's binding site.

2.2. Docking studies

The furin X-ray crystal structure recently published by Henrich, et al.⁷ is valuable to rationalize the binding affinity of known furin inhibitors and for the design of new compounds. The binding

pocket and S subsites of furin are thoroughly described elsewhere.⁷ Highlights and residues forming each of the sub sites are summarized in Table 3. This information is at hand to analyze molecular models of furin inhibitors.

We developed binding models of a set of furin inhibitors, by means of automated molecular docking. Docking studies were focused on furin's catalytic site. The selected molecules for docking studies correspond to peptides that, based on their high similarity to the co-crystallized ligand, are very likely to bind to the catalytic site of the enzyme. In addition, we chose to analyze peptides with physicochemical properties close to drugs in the property space shown in Figure 2, as they are of greater biological relevance and of suitable size for modeling studies.

Table 4 shows the chemical structures of the compounds selected. This set includes peptide derivatives (molecules 1–16)¹⁷ and peptides (molecules 17–26).^{12,38} The peptide derivatives have the general sequence R-Arg-Val-P2-P1, where R = Phac, Dec, Ac, and P2 = Arg, Lys. All of these sequences contain C-terminal non-peptide substituents at the P1 position, which were substituted with different functional groups as shown in Table 4.^{17,19} The entire set contains highly potent furin peptide inhibitors as well as weak binders; experimental K_i values range from 0.81 to 9710 nM. A modified version of the irreversible inhibitor from the furin crystal structure⁷ is also included in the set and corresponds to molecule 27.^{7,44}

2.3. Docking of the co-crystallized ligand

Upon binding of the irreversible inhibitor decanoyl-RVCR-chloromethylketone (dec-RVCR-cmk) with furin, the cmk group makes two covalent bonds with the enzyme. The methylene group binds

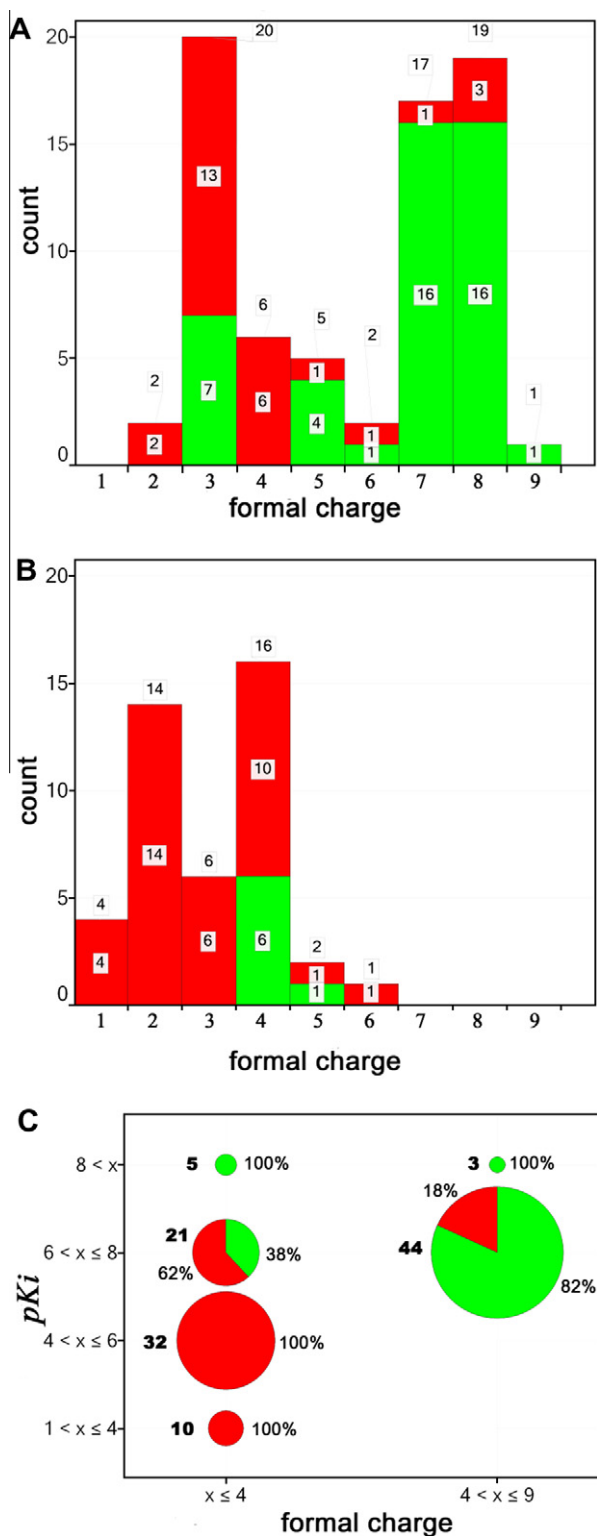


Figure 3. FCharge distribution of active and inactive furin inhibitors (peptides and small molecules) from ChEMBL. (A) FCharge count of peptides. (B) FCharge count of small molecules. (C) Traffic-light pie chart of FCharge integrating PKi. Active compounds are shown in green and inactive in red.

to His194, and the Arg carbonyl group makes a hemiketal bond with Ser368. To mimic the unbound ligand dec-RVKR-cmk for our docking studies, we cleaved the covalent bonds originally found in the X-ray structure of furin, modified the secondary alcohol to a carbonyl and capped the chain with a $-CH_3$ terminal group. The docking model obtained shows that the orientation of each of

the P side chains is similar between our docking model and the co-crystallized ligand (Fig. S1 in Supplementary data) supporting the use of the docking protocol.

Table 5 summarizes the docking scores from Glide XP and the occupancy of subsites of furin protein for the ligands. The ligand efficiency, calculated as the docking score in relation to the number of non-hydrogen atoms is also included. Glide scores ranged from -7.56 to -14.86 kcal/mol. Noteworthy, the most active inhibitor, **13**, has the highest docking score, -14.86 kcal/mol. Docking results show that subsites from S1 to S4 can be occupied by the corresponding side chains P1 to P4 or they can be swapped. It has been reported that S1 and S4 subsites have specificity for arginine in the peptide-protease recognition process.⁴⁵ From our models and from the point of view of inhibitor design, in addition to arginine side chains, S1 and S4 could be occupied by side chains with equivalent electrostatic nature and shape, for example, guanidine substituted benzyl rings. For instance, the relative occupancy of the active ligands **2** and **13** (Table 5) is opposite. The hypothesis of a different binding mode with respect to the co-crystallized ligand requires experimental confirmation. The most active ligand **13** shows occupancy equivalent to that of the co-crystallized ligand **27**, where the positively charged (amidomethyl)benzamidinium moiety (P1) occupies the subsite S1, and P2, P3, and P4 occupy S2, S3 and S4, respectively. However, for ligand **2** P1 moiety occupies subsite S4 and P4 is located in S1. These binding poses are exemplified in Figure 4. Docking analysis of the active 4-amidinobenzylamide compounds **13–16**, suggests that these compounds bind with the same occupancy as the co-crystallized ligand corroborating that 4-amidinobenzylamide moiety could be a surrogate for arginine in P1.¹⁹ In addition, comparison of docking binding modes between analog compounds **14** and **16** suggests that subsite S2 can be occupied by arginine or lysine alike. These observations are in line with the tendency of small molecules to have better affinity to furin binding site for incrementally negatively charged structures.

It is worth mentioning that subsites S2–S4 are shallow and solvent exposed. Consequently, the negatively charged residues forming these subsites can interact with more than one P group and do not necessarily follow the interaction pattern Sn–Pn. These observations are reasonable, for example, a closer look at Table 5 shows that when the interaction pattern Sn–Pn is not followed (i.e., molecule **2**, **4**, **5**, etc.) the phenyl acetate group ($R = Phac$) can be accommodated in the S1' or S2' subsites. Inspection of these binding poses shows that the phenyl ring not only can make $\pi-\sigma$ or $\pi-\pi$ interactions with His194 and His364, but also can make cation- π (salt bridge) interactions with Arg193. In addition, the phenyl group can be accommodated at subsite S2' making favorable aromatic hydrogen bond interactions with the hydroxyl group of Thr365.

Unlike other convertases, like PC2, all furin subsites are negatively charged¹² therefore, residues with side chains positively charged could be accommodated on any Sn subsite, if they have a suitable shape complementary (see Table 3).⁴⁶ It has also been shown that mutation of negatively charged residues alters the specificity of furin.⁴⁷ Thus, as a next step, we focused on electrostatic interactions involved in the furin-inhibitor recognition process.

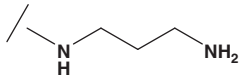
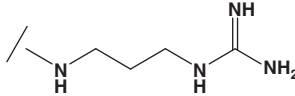
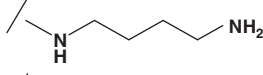
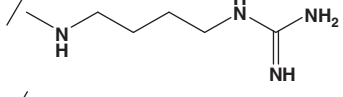
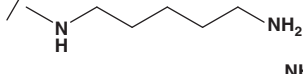
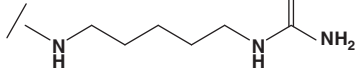
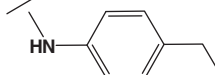
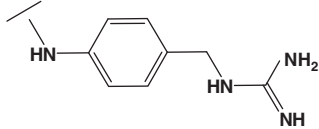
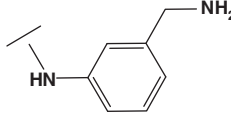
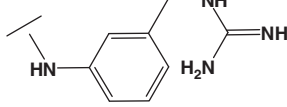
As a post docking procedure, we herein used protein–ligand interaction fingerprints (PLIF). PLIFs are simply strings that convert protein–ligand interactions from 3-D information into 1-D representations. Interactions such as ionic interactions, hydrogen bonds, and surface contacts are categorized according to the residue of origin, and built into a fingerprint scheme which is representative of a given dataset of protein–ligand complexes or poses into the binding site. Each fingerprint bit is denoted by a letter as follows: sidechain hydrogen bond donor (D), sidechain hydrogen bond acceptor (A), backbone hydrogen bond donor (d), backbone hydrogen bond acceptor (b), solvent hydrogen bond (O), Ionic attraction

Table 3
Important residues in the vicinity of furin proteolytic site

Site	Residues forming the site	Functional group preferred in the ligand	Notes
S6	Glu230, Asp233		If P4 is short, P6 could go to S4
S5	Glu257		
S4	Glu236, Asp264, Trp254, Tyr308, Glu257	Arg	P4* sidechain extend inward within the cleft
S3	Glu257	Val	P3* extends outward into the solvent
S2	Asp154, Asn192, Asp191	Lys, Arg	Hydrogen bond interactions
S1	Ser253-Gly255 and Ser293-Asn295, Asp258, Asp306, Ala292, Pro256	Arg	Site formed by hydrophobic groups, interaction with carboxylates and carbonyl groups
Entrance frame	Ser253, Trp254, Gly255, Pro256, Glu257, Asp258		Entrance frame to the S1pocket
Triad	Ser368, His194, Asp153		
Oxyanion hole	Asn295, Ser368		
S1'		No Lys	Accommodates polar or negatively charged P1'
S2'		No Lys	S2' host medium sized hydrophobic residues

* P groups of the co-crystallized ligand.

Table 4
Structure of furin inhibitors used in docking studies

Compound	K _i (nM)	R	P4	P3	P2	P1	Refs.
1	3020	Phac	Arg	Val	Arg		
2	63	Phac	Arg	Val	Arg		
3	7490	Phac	Arg	Val	Arg		
4	78	Phac	Arg	Val	Arg		16
5	553	Phac	Arg	Val	Arg		
6	1070	Phac	Arg	Val	Arg		
7	627	Phac	Arg	Val	Arg		
8	1430	Phac	Arg	Val	Arg		16
9	1320	Phac	Arg	Val	Arg		
10	2730	Phac	Arg	Val	Arg		

(continued on next page)

Table 4 (continued)

Compound	K_i (nM)	R	P4	P3	P2	P1	Refs.
11	9710	Phac	Arg	Val	Arg		16
12	53	Phac	Arg	Val	Arg		
13	0.81	Phac	Arg	Val	Arg		
14	1.6	Dec	Arg	Val	Arg		
15	1	Ac	Arg	Val	Arg		16
16	3.3	Dec	Arg	Val	Lys		
17	1400	Ac-Leu-Leu	Arg	Val	Lys	Arg-NH ₂	17
18	1900	Ac-Leu-Met	Arg	Val	Lys	Arg-NH ₂	
19	190	Ac-Leu-Lys	Arg	Val	Lys	Arg-NH ₂	
20	3400	Ac-Leu-Tyr	Arg	Val	Lys	Arg-NH ₂	
21	800	H-Leu-Leu	Arg	Val	Lys	Arg-NH ₂	10
22	3500	Ac-Leu-Leu	Arg	Val	Lys	Arg-OH	
23	420	H-Leu-Leu	Arg	Val	Lys	Arg-OH	10
24	6400	H-Arg	H-Arg	Arg	Arg	Arg-OH	
25	990	H-Arg	Arg	Arg	Arg	Arg-OH	14
26	114	H-Arg-Arg	Arg	Arg	Arg	Arg-OH	
27	—	Dec	Arg	Val	Lys	Arg-CH ₃	

Molecule 27 resembles the co-crystallized ligand.

Table 5

Glide docking scores and furin inhibitory potency of the ligands studied here. Sub-sites occupancy by furin ligands

Compound	Docking score (kcal/mol)	Ligand efficiency	S5	S4	S3	S2	S1	S1'	S2'
1	−11.37	−0.26	Phac	P4	P3	P2	P1		
2	−12.80	−0.28		P1	P2	P3	P4	Phac	
3	−9.00	−0.20	Phac	P4	P2	P1	—		
4	−12.48	−0.27		P1	P2	P3	P4		Phac
5	−12.84	−0.29		P1	P2	P3	P4	Phac	
6	−10.89	−0.23		P1	P2	P3	P4	Phac	
7	−13.54	−0.28		P1/P2	P2	P3	P4	Phac	
8	−14.32	−0.28		P1	P3	P2	P4	Phac	
9	−11.86	−0.25		P1	P3	P2	P4		Phac
10	−9.64	−0.19		P2	P1/P3	P4	—	Phac	
11	−9.72	−0.21	Phac	P4	P3	P2	P1		
12	−11.72	−0.24		P1	P4	P2	P3		Phac
13	−14.86	−0.30	Phac	P4	P3	P2	P1		
14	−12.26	−0.24		P4	P3	P2	P1		
15	−10.45	−0.24	Ac	P4/P3	P2	—	P1		
16	−11.24	−0.23		P4	P3	P2	P1		
17	−14.66	−0.25		P4	P3	P2	P1		
18	−10.13	−0.17		P4/P3	P2	P1			
19	−11.07	−0.19		P3	P2/P4	P1			
20	−11.90	−0.19		P2/P1	P3/P1	—	P4	Tyr	Leu-Ac
21	−12.37	−0.22		P4	P3	P2	P1		
22	−14.32	−0.25		P4	P3	P2	P1		
23	−13.42	−0.24		P4	P3	P2	P1		
24	−12.46	−0.28		P4	P1	P3	P2		
25	−13.48	−0.24		P4	P3	P2	P1		
26	−7.56	−0.11		P1/P3	P3	P2/P4	P5		
27	−8.01	−0.16		P4	P3	P2	P1		

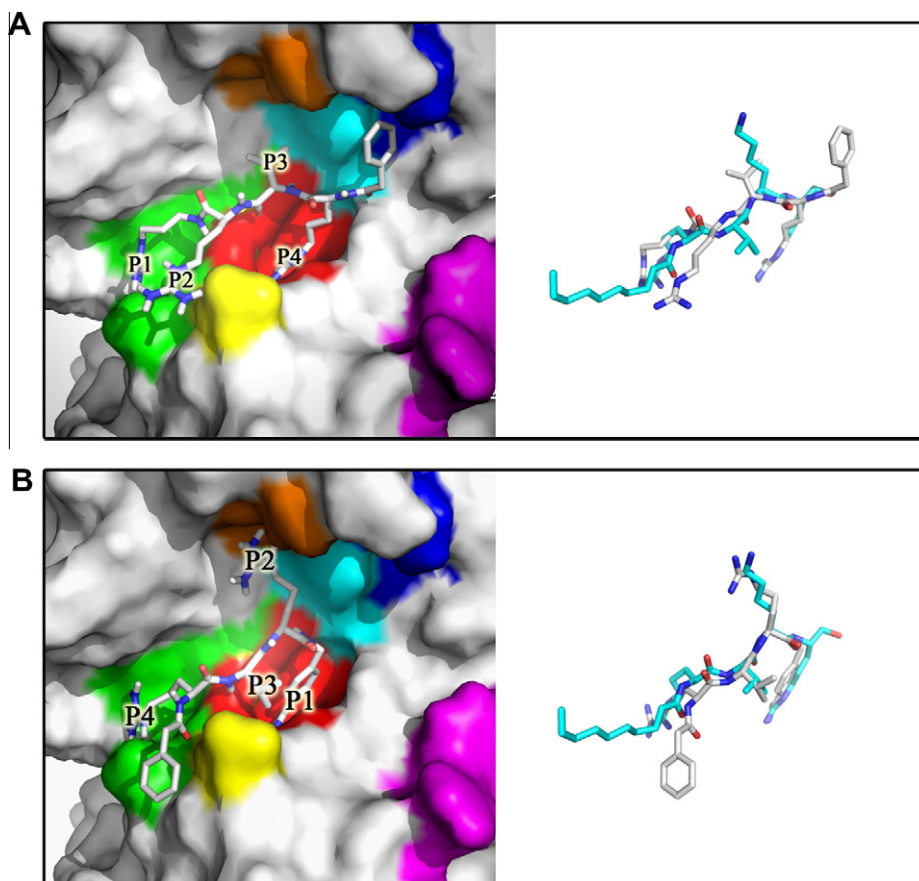


Figure 4. Relevant docking poses representing the two possible binding modes of peptides. Binding mode of ligands **2** (A) and **13** (B) showing opposite and similar orientations with respect to the co-crystallized ligand, respectively.

(I), and surface contact (C). In this study we focused our analysis on hydrogen bonds and ionic interactions.

Figure 5 depicts the barcode representation and counts for the docking interactions between furin inhibitors and the negatively charged residues Asp and Glu from the subsites S1–S4. Figure 5A shows that electrostatic interactions (barcodes in black) predominate on S1 (Asp258 and Asp306), S2 (Asp154 and Asp191), and S4 (Glu236 and Asp264). It was observed that the S3 subsite was frequently occupied by valine; as a result, Glu257 participates on much less ionic interactions compared to S1, S2, and S4 subsites. The frequency count of each interaction is shown in Figure 5B. It is evident that Glu257 (S3) and Asp191 (S2) participated in few interactions.

Inspection of the binding interactions of active and inactive furin inhibitors in Figure 5 does not provide a clear explanation of the difference in activity, suggesting that further refinement of the models might be required or that the slight differences in binding interactions have a large impact in the biological activity. This behavior is frequently the result of high structural similarity of the ligands. Similar molecules that have different biological activity are known as activity cliffs.⁴⁸ As noted before, activity cliffs can be valuable to identify pharmacophoric features, but are often elusive by molecular modeling studies.^{49,50} The presence of activity cliffs in this set highlights the need of exploring structurally related furin inhibitors, such as those found on combinatorial libraries.

3. Conclusions

Analysis of the experimental inhibition constants, K_i , of the dataset studied here indicates that furin inhibitors with peptide nature provide potent binders more consistently than small molecule-based inhibitors. In addition, although finding active peptides

with physicochemical properties closer to those of known drugs is more challenging, peptides in that region of the chemical space can be even more potent and are therefore more amenable as lead compounds compared to peptides of larger size and higher content of positive formal charges.

The two putative binding modes obtained in our docking studies are reasonable; the experimental confirmation of the alternative binding mode is required. Relevant structural features derived from these binding modes are: (a) the S1 subsite hosted a benzimidamide group, (larger, shorter or more flexible side chains are less favorable in this site); and (b) S1 and S4 subsites participated regularly on hydrogen bond and ionic interactions, followed by S2, but these interactions were seldom observed on S3.

Taken together, analyses of furin inhibitors in terms of physicochemical properties and docking studies show the importance of the formal charge, the need of exploring highly similar structures, and the difficulty but feasibility of arriving at potent peptide-based inhibitors with physicochemical properties similar to known drugs.

4. Methods

4.1. Chemoinformatic analysis of furin inhibitors from ChEMBL

A total of 356 structures were retrieved from the ChEMBL database (downloaded January 2012, from <http://www.ebi.ac.uk/chembl/db/>), entering 'furin' as a search criterion, and was submitted by the option 'assays'. All structures were then processed with the MOE package. Energy minimization was performed using the MMFF94x force field; with default parameters in MOE. The final set contains 115 unique compounds with accurately estimated inhibition constants (K_i). The sum of formal charges (FCharge)

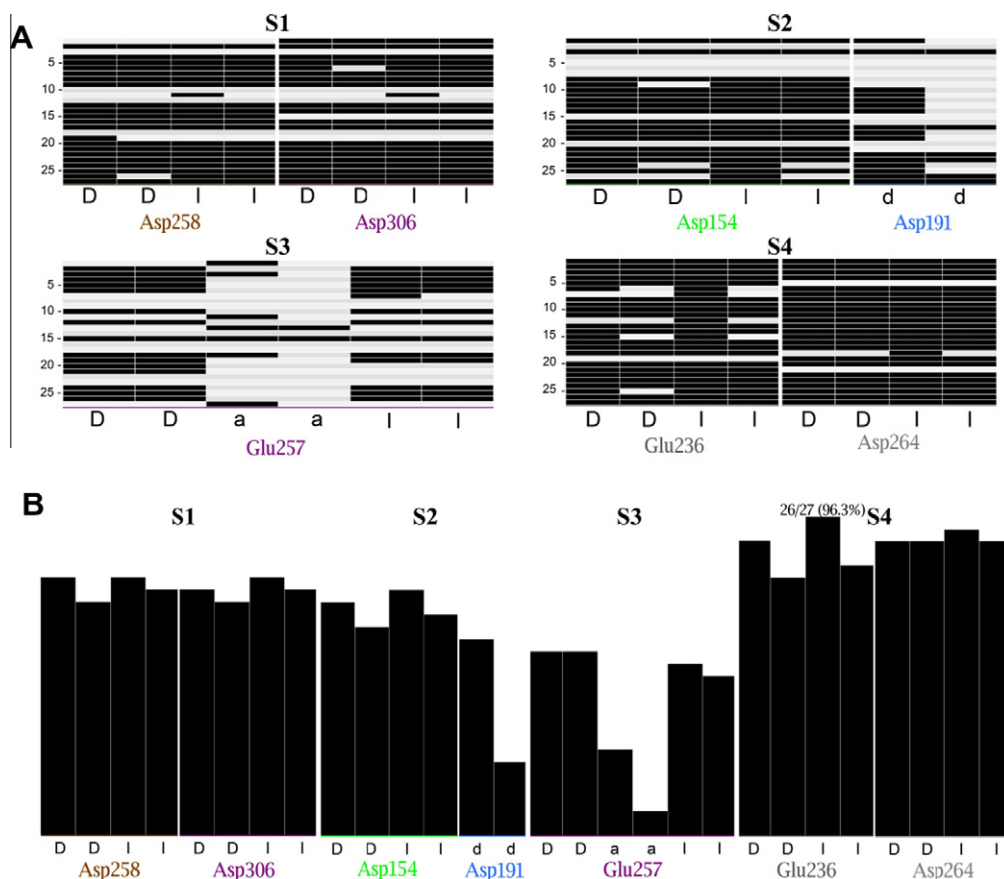


Figure 5. Barcode representation (A) and population (B) of the docking interactions between furin inhibitors and the ionic residues Asp and Glu from the subsites S1, S2 S3, and S4. The numbers on the left-side of the barcode representation identify the ligands 1–27. Each fingerprint bit is denoted by a character to indicate its meaning: sidechain hydrogen bond donor (D), backbone hydrogen bond donor (d), backbone hydrogen bond acceptor (a), and Ionic attraction (I).

was calculated for all compounds using MOE. The property space representation of the selected furin inhibitors was built by principal components analysis (PCA) using Spotfire 9.1.1. software.⁵¹ Known drugs were included, consisting of 1731 compounds obtained from DrugBank and implemented in the ZINC⁵² database, downloaded in March 2011. PCA was carried out considering the six scaled molecular properties: molecular weight (MW), number of rotatable bonds (RB), the octanol/water partition coefficient (SlogP), topological polar surface area (TPSA), hydrogen bond acceptors (HBA), and hydrogen bond donors (HBD). To visualize the chemical space, the first two principal components were plotted. Pie charts, and counts were also performed in Spotfire.

4.2. Molecular modeling of furin inhibitors

A dataset containing 26 molecules with known inhibitory potency towards furin was collected from the ChEMBL database and from the literature.^{7,12,17,19,38} Chemical structures are shown in Table 3. The structures were built using Maestro version 9.1.⁵³ To generate an initial conformation of the molecules for the docking studies, a conformational ensemble of each molecule, generated with OMEGA, was aligned to the co-crystallized ligand using ROCS. LigPrep was then utilized to produce different protomers for each ligand.

4.3. Docking protocol

The atomic coordinates of furin were obtained from the RCSB Protein Databank⁵⁴ (PDB code: 1P8J). Enzyme preparation was

performed using the Protein Preparation Wizard implemented in Maestro.⁵³ The dimensions of the inner and outer grids (in Å) were $14 \times 14 \times 14$ and $55 \times 55 \times 55$, respectively. The inhibitors were docked utilizing Glide version 5.5, the binding poses were evaluated utilizing Glide extra precision (SP followed by XP and XP refine) scoring function.⁵⁵

4.4. Protein ligand interaction fingerprints (PLIF)

PLIF is a tool for summarizing the interactions between ligands and receptors using a fingerprint scheme. PLIF is implemented in MOE package software (Molecular Operating Environment, version 2010.10). In this work, the binding poses obtained from Glide XP refine docking were used as input for PLIF with default settings.

Acknowledgments

K.M.M. thanks the State of Florida, Executive Officer of the Governor's Office of Tourism, Trade and Economic Development for funding. Authors thank OpenEye for providing the programs ROCS, OMEGA and VIDA, J. Medina and I. Lindberg for helpful discussions, T. Caulfield for participating on an early version of this work and J. Waddell for assistance to analyse the peptide dataset and proofreading.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bmc.2012.05.029>.

References and notes

- Seidah, N. G.; Chretien, M. *Brain Res.* **1999**, 848, 45.
- Steiner, D. F. *Curr. Opin. Chem. Biol.* **1998**, 2, 31.
- Roebroek, A. J. M.; Umans, L.; Pauli, I. G. L.; Robertson, E. J.; van Leuven, F.; Van de Ven, W. J. M.; Constam, D. B. *Development* **1998**, 125, 4863.
- Rouille, Y.; Bianchi, M.; Irminger, J. C.; Halban, P. A. *FEBS Lett.* **1997**, 413, 119.
- Bennett, B. D.; Denis, P.; Haniu, M.; Teplow, D. B.; Kahn, S.; Louis, J. C.; Citron, M.; Vassar, R. *J. Biol. Chem.* **2000**, 275, 37712.
- Coppola, J.; Bhojani, M. S.; Ross, B. D.; Rehemtulla, A. *Neoplasia* **2008**, 10, 363.
- Henrich, S.; Cameron, A.; Bourenkov, G. P.; Kiefersauer, R.; Huber, R.; Lindberg, I.; Bode, W.; Than, M. E. *Nat. Struct. Biol.* **2003**, 10, 669.
- Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, 27, 157.
- Henrich, S.; Lindberg, I.; Bode, W.; Than, M. E. *J. Mol. Biol.* **2005**, 345, 211.
- Vivoli, M.; Caulfield, T. R.; Martínez-Mayorga, K.; Johnson, A. T.; Jiao, G. S.; Lindberg, I. *Mol. Pharmacol.* **2012**, 81, 440.
- Fugere, M.; Day, R. *Trends Pharmacol. Sci.* **2005**, 26, 294.
- Cameron, A.; Appel, J.; Houghten, R. A.; Lindberg, I. *J. Biol. Chem.* **2000**, 275, 36741.
- Antopolsky, M.; Basak, A.; Khatib, A.-M.; Mohottalage, D.; Basak, S.; Kolajova, M.; Bag, S. S.; Basak, A. *PLoS One* **2009**, 4, e7700.
- Kacprzak, M. M.; Peinado, J. R.; Than, M. E.; Appel, J.; Henrich, S.; Lipkind, G.; Houghten, R. A.; Bode, W.; Lindberg, I. *J. Biol. Chem.* **2004**, 279, 36788.
- Basak, A. *J. Mol. Med. (Heidelberg, Germany)* **2005**, 83, 844.
- Bontemps, Y.; Scamuffa, N.; Calvo, F.; Khatib, A.-M. *Med. Res. Rev.* **2007**, 27, 631.
- Fugere, M.; Appel, J.; Houghten, R. A.; Lindberg, I.; Day, R. *Mol. Pharmacol.* **2007**, 71, 323.
- López-Vallejo, F.; Giulianotti, M. A.; Houghten, R. A.; Medina-Franco, J. L. *Drug Discovery Today* **2012**, <http://dx.doi.org/10.1016/j.drudis.2012.04.001>, in press.
- Becker, G. L.; Sielaff, F.; Than, M. E.; Lindberg, I.; Routhier, S.; Day, R.; Lu, Y.; Garten, W.; Steinmetzer, T. *J. Med. Chem.* **2010**, 43, 1067.
- Yongye, A. B.; Li, Y. M.; Giulianotti, M. A.; Yu, Y. P.; Houghten, R. A.; Martínez-Mayorga, K. *J. Comput. Aided Mol. Des.* **2009**, 23, 677.
- Li, Y.; Yongye, A.; Giulianotti, M.; Martínez-Mayorga, K.; Yu, Y.; Houghten, R. A. *J. Comb. Chem.* **2009**, 11, 1066.
- Rodríguez, R. A.; Pan, P.-S.; Vasko, R. C.; Pan, C.-M.; Disman, W. S.; McAlpine, S. R. *J. Mex. Chem. Soc.* **2008**, 52, 201.
- Yongye, A. B.; Pinilla, C.; Medina-Franco, J. L.; Giulianotti, M. A.; Dooley, C. T.; Appel, J.; Nefzi, A.; Scior, T.; Houghten, R.; Martínez-Mayorga, K. *J. Mol. Model.* **2011**, 17, 1473.
- López-Vallejo, F.; Caulfield, T.; Martínez-Mayorga, K.; Giulianotti, M. A.; Nefzi, A.; Houghten, R. A.; Medina-Franco, J. L. *Comb. Chem. High Throughput Screening* **2011**, 14, 475.
- Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martínez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agraftotis, D. K. *J. Chem. Inf. Model.* **2012**, 52, 867.
- Scior, T.; Medina-Franco, J. L.; Do, Q. T.; Martínez-Mayorga, K.; Rojas, J. A. Y.; Bernard, P. *Curr. Med. Chem.* **2009**, 16, 4297.
- Kolb, P.; Irwin, J. J. *Curr. Top. Med. Chem.* **2009**, 9, 755.
- Shoichet, B. K.; McGovern, S. L.; Wei, B. Q.; Irwin, J. J. *Curr. Opin. Chem. Biol.* **2002**, 6, 439.
- Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. *J. Med. Chem.* **2006**, 49, 5912.
- Kroemer, R. T. *Curr. Protein Pept. Sci.* **2007**, 8, 312.
- Bender, A. *Expert. Opin. Drug Discov.* **2010**, 5, 1141.
- Medina-Franco, J. L.; Martínez-Mayorga, K.; Giulianotti, M. A.; Houghten, R. A.; Pinilla, C. *Curr. Comput. Aided Drug Des.* **2008**, 4, 322.
- Varnek, A.; Baskin, I. *Mol. Inf.* **2011**, 30, 20.
- Xue, L.; Bajorath, J. *Comb. Chem. High Throughput Screening* **2000**, 3, 363.
- Xu, J.; Hagler, A. *Molecules* **2002**, 7, 566.
- Anglikar, H. J. *Med. Chem.* **1995**, 38, 4014.
- Shiryayev, S. A.; Remacle, A. G.; Ratnikov, B. I.; Nelson, N. A.; Savinov, A. Y.; Wei, G.; Bottini, M.; Rega, M. F.; Parent, A.; Desjardins, R.; Fugere, M.; Day, R.; Sabet, M.; Pellicchia, M.; Liddington, R. C.; Smith, J. W.; Mustelin, T.; Guiney, D. G.; Lebl, M.; Strongin, A. Y. *J. Biol. Chem.* **2007**, 282, 20847.
- Apletalina, E.; Appel, J.; Lamango, N. S.; Houghten, R. A.; Lindberg, I. *J. Biol. Chem.* **1998**, 273, 26589.
- Worachartcheewan, A.; Nantasenamat, C.; Naenna, T.; Isarankura-Na-Ayudhya, C.; Prachayasittikul, V. *Eur. J. Med. Chem.* **2009**, 44, 1664.
- Sielaff, F.; Than, M. E.; Bevec, D.; Lindberg, I.; Steinmetzer, T. *Bioorg. Med. Chem. Lett.* **2011**, 21, 836.
- Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery. Rev.* **1997**, 23, 3.
- Oprea, T. I. *J. Comput. Aided Mol. Des.* **2002**, 16, 325.
- Blake, J. F. *Biotechniques* **2003**, 16.
- Jean, F.; Stella, K.; Thomas, L.; Liu, G.; Xiang, Y.; Reason, A. J.; Thomas, G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95, 7293.
- Izidoro, M. A.; Gouvea, I. E.; Santos, J. A. N.; Assis, D. M.; Oliveira, V.; Judice, W. A. S.; Juliano, M. A.; Lindberg, I.; Juliano, L. *Arch. Biochem. Biophys.* **2009**, 487, 105.
- Nakayama, K. *Biochem. J.* **1997**, 327, 625.
- Creemers, J.; Siezen, R.; Roebroek, A.; Ayoubi, T.; Huylebroeck, D.; Van de Ven, W. *J. Biol. Chem.* **1993**, 268, 21826.
- Johnson, M. A.; Maggiora, G. M. *Concepts and Applications of Molecular Similarity*; New York John Wiley & Sons, 1990.
- Maggiora, G. M. *J. Chem. Inf. Model.* **2006**, 46, 1535.
- Yongye, A. B.; Byler, K.; Santos, R.; Martínez-Mayorga, K.; Maggiora, G. M.; Medina-Franco, J. L. *J. Chem. Inf. Model.* **2011**, 51, 1259.
- Spotfire, version 9.1.1, TIBCO Software, Inc., Somerville, MA. Available at: <http://spotfire.tibco.com>.
- Irwin, J. J.; Shoichet, B. K. *J. Chem. Inf. Model.* **2005**, 45, 177.
- MAESTRO 9.2; Schrödinger, LLC: New York, NY, 2011.
- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, 28, 235.
- GLIDE 5.5; Schrödinger, LLC: New York, NY, 2009.