A group of presumed drug-like molecules that possess high in silico affinity for angiotensin-converting enzyme 2 were computationally designed. This enzyme is a promising new target in both cardiorenal disease and some coronavirus infections. A set of substrate analogous molecules were optimized by means of the LeapFrog module of the SYBYL package. Later, Molinspiration and Molsoft were used for screening out the compounds with low oral bioavailability. Similarly, OSIRIS was used for screening out the compounds having serious side effects. At the end of several stages of screening, seven candidates to anti-viral drugs fulfilling all the evaluated criteria were obtained. They are amenable for future studies in vitro and in vivo. These designed ligands were finally evaluated by Quantitative Structure Activity Relationship studies. 21 molecules were used to carry out the qsar models. From these four molecules were taken as external sets yielding models with $q^2 = 0.652$ and $r^2 = 0.962$ values.

Keywords: Angiotensin; SYBYL; QSAR; Cheminformatics.

INTRODUCTION

The monocarboxypeptidase angiotensin-converting enzyme 2 (ACE2) is a novel biochemical component of the renin–angiotensin system whose discovery has added a further layer of complexity to the classical concept of this cardiovascular regulatory system. Hitherto, its physiological and pathological roles are not fully deciphered but it has been reported that ACE2 is implicated in cardiovascular and renal biology, obesity, inflammatory activity, and diabetes. This enzyme has also been established as the functional receptor of both SCV (the etiological agent of the Severe Acute Respiratory Syndrome) and H-CoV NL63 virus. Hereby, ACE2 is considered an important target in the control of those diseases.

Developing anti-SCV therapeutical agents is of paramount importance due to the high lethality of SCV infections, its enormous economic and social impact, fears of renewed outbreaks as well as the potential misuse of such viruses as biologic weapons.

The monocarboxypeptidase angiotensin-converting enzyme 2 (ACE2) is a novel biochemical component of the renin–angiotensin system whose discovery has added a further layer of complexity to the classical concept of this cardiovascular regulatory system. Hitherto, its physiological and pathological roles are not fully deciphered but it has been reported that ACE2 is implicated in cardiovascular and renal biology, obesity, inflammatory activity, and diabetes. This enzyme has also been established as the functional receptor of both SCV (the etiological agent of the Severe Acute Respiratory Syndrome) and H-CoV NL63 virus. Hereby, ACE2 is considered an important target in the control of those diseases.

Developing anti-SCV therapeutical agents is of paramount importance due to the high lethality of SCV infections, its enormous economic and social impact, fears of renewed outbreaks as well as the potential misuse of such viruses as biologic weapons. Besides, it is known that diverse coronaviruses bind to a similar region of ACE2 implying that ACE2 ligands could inhibit both infections by SCV and HCoV-NL63. Although the HCoV-NL63 virus circulating in the human population is only modestly pathogenic in most cases, an inhibitor would be useful in the treatment of a subset of serious cases or more pathogenic forms of the virus. In this work computational methods were applied for obtaining inhibitors of the viral fusion event as they have previously proven their ability of targeting ACE2.

PREPARATION OF THE STRUCTURES, CONFORMATIONAL SEARCH AND CONSENSUS

The 1R4L structure corresponding to an X-ray crystallography of the extracellular metallopeptidase domain of ACE2 in complex with the MLN-4760 ligand was extracted from the Brook Haven PDB. The ACE2 inhibitors developed by Dales et al. were sketched, energetically minimized, and charged in the SYBYL package version 7.0. In addition, the optimized structures of these compounds were prepared for running docking experiments following the same procedure. FlexX program was used as search algorithm for finding 100 binding poses of the ligands with respect to the structure of enzyme 1R4L obtained from the RCSB Protein Data Bank. In the definition of the active site for the docking simulations a radius of 7.2 Å from His was taken. Subsequently, one of the 100 poses was selected through rank by rank consensus among the scoring functions G score, PMF Score, D score and Chemscore. The same docking protocol was implemented for each study of Quantitative Structure Activity...
Relationship (QSAR).

**QSAR STUDIES**

A set of 21 compounds [Figure 1, Table 1] taken from a publication of Dales et al. was utilized for the 3D-QSAR CoMSIA (Comparative Molecular Similarity Analysis) analyses. These models were developed in the SYBYL 7.0 package using the ligands in the Table 1 in their top ranked pose after docking in FlexX. No further alignment of the ligands was performed. CoMSIA descriptors (steric, electrostatic, hydrophobic, hydrogen bond donor, hydrogen bond acceptor) were calculated with the standard Tripos force field at every point of the three dimensional lattice, using the sp3 carbon probe with +1 charge with standard CoMSIA cut-off values.

The method of partial least squares (PLS) implemented in the QSAR module of SYBYL was used to construct and validate the models. Column filtering was set to 2.0 kcal/mol to speed up the analysis and reduce the noise. The CoMSIA descriptors served as independent variables and pIC50 values as dependent variable in PLS regression analysis. The performance of the derived models was calculated using the leave one out (LOO) cross-validation method. The optimum number of components (Nc) used to derive the non cross-validated model was defined as the number of components leading to the highest r2 cross-validated (q2) and lowest standard error of prediction (SEP).

To obtain the statistical confidence limits on the analyses, bootstrapping was carried out with 100 groups.

**LEAD OPTIMIZATION**

Optimization was carried out by LeapFrog (LF). LF is a second-generation lead optimization tool to design a series of potential active ligand molecules. Binding energy calculations in LF are performed by three major components: direct steric, electrostatic, and implicit hydrogen bonding enthalpies of ligand–cavity binding using the Tripos force field. In LF, ligand atom coordinates are binned to increase speed and the binding energy of each ligand atom is calculated as though the atom were actually located.

**Table 1. Biological Activity Values of the Compounds Reported by Dales et al.**

| R1 | R2 | R3 | IC50 |
|----|----|----|------|
| 1RS* | — | CH(CH3)2 | H | 1.4 |
| 1SR | — | CH(CH3)2 | H | 2.2 |
| 1SS | — | CH(CH3)2 | H | 1.2 |
| 2 | — | CH(CH3)2 | Benzyl | 10 |
| 3 | Benzyl | CH(CH3)2 | — | 0.024 |
| 4 | Benzyl | Me | — | 0.30 |
| 5 | Benzyl | Ph | — | 0.34 |
| 6 | Ciclohexyl CH2 | CH(CH3)2 | — | 0.27 |
| 7* | Ciclohexyl (CH3)2CH2 | CH(CH3)2 | — | 0.10 |
| 8 | 4-NO2Benzyl | CH(CH3)2 | — | 0.076 |
| 9 | 4-ClBenzyl | CH(CH3)2 | — | 0.021 |
| 10 | 4-CF3Benzyl | CH(CH3)2 | — | 0.052 |
| 11 | 4-MeBenzyl | CH(CH3)2 | — | 0.032 |
| 12* | 2-MeBenzyl | CH(CH3)2 | — | 0.29 |
| 13 | 3-MeBenzyl | CH(CH3)2 | — | 0.0042 |
| 14 | 3,4-diMeBenzyl | CH(CH3)2 | — | 0.010 |
| 15 | 3,5-diMeBenzyl | CH(CH3)2 | — | 0.0014 |
| 16RR | 3,5-diClBenzyl | CH(CH3)2 | — | 0.072 |
| 16RS | 3,5-diClBenzyl | CH(CH3)2 | — | 0.0014 |
| 16RS* | 3,5-diClBenzyl | CH(CH3)2 | — | 0.470 |
| 16SS | 3,5-diClBenzyl | CH(CH3)2 | — | 0.00044 |

* Molecule of the QSAR test set
in the center of a cube containing that atom. A simple linear expression then yields the energy of interaction between the site and that particular ligand atom. Summing over all ligand atoms yields the overall site–ligand interaction energy. When the LF program was executed, the bin created contained the 22 closest residues to the co-crystallized ligand in 1R4L. This cavity was used to generate the Site-points. The charge of a Site-point atom is positive, negative, or lipophilic. No optional cavity desolvation energy was used. On each of the inhibitors from Table 1 ten thousand modifications were stochastically performed.

VIRTUAL SCREENING

In silico ADMET has taken an increasingly significant place within the drug development pipeline. The structural optimization process in LF gives rise to a profusion of chemical structures with high ligand efficiency. To assess their real pharmaceutical usefulness successive filtering by means of on-line services\textsuperscript{23} were implemented as depicted in Figure 2. A set of QSAR-based proposed candidates underwent screening as well. The employed tools for prediction of drug-likeness were Molinspiration [http://www.molinspiration.com/cgi-bin/properties], OSIRIS\textsuperscript{24} and Molsoft [http://www.molsoft.com/mprop/]. Drug-likeness may be defined as a complex balance of various molecular properties and structure features which determine whether a particular molecule is similar to the known drugs.

A widely used filter for drug-like properties is known as Lipinski’s rule of five (RO5). The original RO5 deals with orally active compounds and defines four simple physicochemical parameter ranges (\text{MWT} \leq 500, \text{log P} \leq 5, \text{H-bond donors} \leq 5, \text{H-bond acceptors} \leq 10) associated with 90% of orally active drugs that have achieved phase II clinical status.\textsuperscript{25} On the other hand, toxicity is responsible for many compounds failing to reach the market and for the withdrawal of a significant number of compounds from the market once they have been approved.\textsuperscript{26}

RESULTS AND DISCUSSION

Docking

The definition of the active site focused most of the computational resources on the hydrophobic core of the enzyme. The deeply recessed and shielded active site of ACE2 is a common structural feature of proteases serving to avoid proteolysis of large structured peptides.\textsuperscript{27} The blind docking performed by eHiTS 6.2\textsuperscript{28} scanned the entire protein rigorously confirming that the binding site of all ligands with the ACE2 protein effectively is the one selected for the docking through FlexX. Inside the specific recognition site where the peptide bond is cleaved were found three pockets which are displayed rendered as a Connolly surface in Figure 2. The hydrophobic S\textsubscript{1} subsite is composed by the residues N\textsuperscript{51}, W\textsuperscript{349}, A\textsuperscript{348}, T\textsuperscript{347}, H\textsuperscript{374}, H\textsuperscript{378}, E\textsuperscript{402}, and Y\textsuperscript{510}. Fig. 3 shows that the pseudo-Leu isobutyl chain of the inhibitors from Table 1 packs nicely into it\textsuperscript{15} but is big enough to accommodate longer chains. In contrast, electrostatic chemical environment of the S\textsubscript{1}' subsite favours interaction with polar residues of the ACE2 natural substrates. This thin pocket is one of the mayor determinants of the substrate specificity of ACE2. In fact, topology of S\textsubscript{1}' in ACE2 differs from that of its homologue in ACE.\textsuperscript{29} S\textsubscript{1}' is composed by the side chain of the residues S\textsuperscript{128}, A\textsuperscript{342},

Fig. 2. Procedure of Successive Filters Used in this Study. Five Criteria Were Taken Into Account: Target Affinity, LF Energy, Lipinski Parameters, Toxicity and Drug-likeness.

Fig. 3. Consensus Pose of a Compound from Table 1 Docked in the Active Site of ACE2. The Carboxicate Groups are located Closed to the Zn Atom Which is Displayed as a Red Sphere. The pseudo-Leu isobutyl chain of the inhibitors is oriented towards the S1 Pocked.
V^{343}, C^{344}, H^{345}, L^{359}, M^{360}, T^{371}, E^{375}, F^{504}, H^{505}, L^{503}, and disulfide link C^{334}/C^{361} which restrict size of the substrate P₁ side chains while suitably are adapted to accommodate the imidazol ring and one of the carboxylate group of the Dales ligands. Further, bulky N-3 substituents at the imidazole ring interact with a third cavity in the ACE2 beyond S'₁ that could be termed S'₂ according to the interaction model of Schechter and Berger²⁰ since it binds to functional substrates. This third pocket is constituted by the residues E^{145}, L^{144}, N^{149}, W^{271}, and R^{273}. The docking study also revealed that inhibitors with stereochemistry 1SR and 16RR exhibit an inverse mode of binding by lodging its benzyl moiety inside the S₁ subsite whereas inhibitor 7 binds such a way that the whole structure of inhibitor 2 is excluded from S₁ subcavity.

**LF Calculations**

It was generated a large database by means of LF, a program for optimization of the ligands docked into their target. In most LF simulations the structures retained their place inside the cavity as the automatic optimization proceeded. The most difficult calculation was that of ligand 7 which only produced 2 LF molecules. Altogether 2372 acceptable structures were obtained. From this database, the structures with the highest LF affinity score were selected. Although any atom in the starting strutures was protected from modifications their imidazole rings and amino dicarboxylate cores were conserved except in the case of the two first molecules from Table 1.

The binding mode of one of the LF generated structures in the active site of ACE2 is displayed in Figure 4. It shares the two carboxylate groups of the compounds from Table 1. These groups coordinate to Zn but the linker group is a methylene instead of an amino group. In addition, the chain in S₁ is enlarged and possesses two fluoride atoms and a ceto group to establish electrostatic interactions with the residues N^{51} and E^{502}. Finally, the ligand E has an oxygen atom that connects the imidazol ring to an ethyl-pyridazine moiety which parcially occupy the S₂ subsite. A pyridazine ring is found within the structure of several pharmaceutical drugs.³¹ Nitrogen atom containing heterocyclic compounds, pyridazines, pyridazinones and phthalazines are important structural feature of many biologically active compounds and show diverse pharmacological properties. Overall, the LF-generated scaffolds fit with high complementarity into the ACE2 catalytic site.

**QSAR Calculations**

In addition, CoMSIA QSAR 3D models were built. The docked-based alignment method led to several significative CoMSIA models in contrast to the CoMFA model and those ones elaborated from the ALIGMENT SYBYL module. To select the best CoMSIA model the possible combinations of their 5 physicochemical standard properties were tried. The respective statistics of the models having $q^2$ values above 0.5 are summarized in the Table 2. The best CoMSIA model from the Dales ligands yielded $q^2 = 0.652$ and $r^2 = 0.962$ values. The best CoMSIA model from the ligands of Table 2 yielded $q^2 = 0.566$ and $r^2 = 0.992$ values. Steric parameter alone gave the best statistics in both cases.

The steric contribution contour maps of CoMSIA are plotted in Figure 5. It shows that the greater potency of ligand 16 (which is also named MLN-4760) with respect to ligand 2 is related to the overlapping of the ligand 16 moieties with the green fields. In fact, the benzene ring of the ligand 16 is farer from the yellow polyhedral than ligand 2.
and the chloride atoms at this ring of the ligand 16 goes further into the green region. In addition, the imidazole ring of ligand 16 changes its orientation so that the N1 atom is much more separated from the unfavourable yellow region of the steric field.

The biological activity of the LF-generated chemo-types were calculated by the CoMSIA models. The results shown in Table 3 suggest that they are more potent than previously reported inhibitors.

ADMET Predictions and Drug-likeness Predictions

The files generated by LF were converted to SMILES format by the Cactus translator (http://cactus.nci.nih.gov/services/translate/). Next, these molecules were submitted to the on-line server MOLINSPIRATION to predict important molecular properties (logP, polar surface area, number of hydrogen bond donors and acceptors, molecular weight, number of Oxygen and Nitrogen atoms, number of Hydrogens attached to N and O, number of violations, number of rotatable bonds and volume). After screening the top 77 LF compounds generated by structural optimization a 50% of hits against the Lipinski’s rule was observed. Prediction of toxicity before the synthesis of compounds ensures the removal of compounds with potential toxic effects. A number of in silico systems for toxicity prediction are available, which help in the classification of toxic and non-toxic compounds. Here, assessment of the studied molecules was performed by means of the OSIRIS property explorer yielding 12 selected prototypes. Since any member of a precomputed set of structural fragment characteristic of harmful compounds was encountered in the screened structures it can be claimed that they are as free of toxic effects as traded drugs. It therefore indicates that there is neither risk of mutagenicity, tumorigenicity, irritability nor reproductive effects. As positive controls were used Ritonavir, Ribavirin, Amantadine and Acyclovir. The fragments in the submitted molecules that gave rise to toxicity alerts are shown in Table 6. According to OSIRIS the most potent compound of Table 1, the MLN4760 co-crystallized ligand, turned out also to be mutagenic by virtue of a benzene ring moiety bearing two chlorides in meta position. Nonetheless, a compound containing a halogenated pyridine ring passed the OSIRIS test.

In addition, Molinspiration server was used to predict the bioactivity over the four most important drug targets: GPCR ligands, kinase inhibitors, ion channel modulators, and nuclear receptors. The method implemented in Molinspiration uses sophisticated Bayesian statistics to compare structures of representative ligands active on the particular target with structures of inactive molecules and to identify substructure features typical for active molecules. Descrip-

| Field       | Statistics Parameters of the QSAR |
|-------------|-----------------------------------|
|             | $q^2$ | $r^2$ | NC | SEE | F value | $R^2_{cv}$ | SD |
| S, E, D     | 0.506 | 0.940 | 6  | 0.214 | 83.952   | 0.010      |
| S, A        | 0.5   | 1.000 | 6  | 0.008 | 24808.95 | 1.000      |
| H           | 0.506 | 1.000 | 10 | 0.014 | 7779.978 | 1.000      |
| S           | 0.652 | 0.962 | 6  | 0.272 | 59.759   | 0.988      |

a Steric field
b Electrostatic field
c Hydrophobic field
d Donor hydrogen bond field
e Acceptor hydrogen bond field
f Non cross-validated square of correlation coefficient
g Number of components
h Standard error of prediction
i From 100 bootstrapping runs
j Standard deviation

Table 3. Predicted Biological Activity of the LF-Generated Structures

| Molecule | pIC$_{50}$ prediction by means of QSAR |
|----------|---------------------------------------|
| A        | 3.46                                  |
| B        | 3.71                                  |
| C        | 3.64                                  |
| D        | 3.52                                  |
| E        | 3.99                                  |
| F        | 3.90                                  |
| G        | 4.00                                  |
tors calculated for the eight previously screened entities are registered in Table 4. As positive controls were included the drugs: propanolol, xylocaine, Imatinib and estradiol. They all are conspicuous drugs targeting the above mentioned classes of receptors. Cellular toxicity was not observed in the molecules ACE2 ligands shown in Table 4, whereas bioactivity was found in every control molecule.

The drug-likeness score was also calculated by the On-line server Molsoft [http://www.molsoft.com/mprop/].
The Molsoft calculated parameters molecular weight, hydrogen bond acceptors and donors, logP, logS, PSA, molecular volume and number of estereogenic centers are registered in Table 5 and Figure 6.

CONCLUSIONS

Since any member of a precomputed set of structural fragment characteristic of harmful compounds was encountered in the screened structures it can be claimed that they are as free of toxic effects as traded drugs. It therefore indicates that there would be no risk of mutagenicity, tumorigenicity, irritability nor reproductive effects. Besides, according to the QSAR-3D models predictions, the LF-designed molecules are more potent than the starting compounds from which they were developed and they could be active in the nanomolar range. Previously other studies had achieved increased the potency of bioactive molecules.33-34

In summary, seven therapeutical entities having high calculated drug-likeliness are proposed to be synthesized and experimentally evaluated. An obstacle regarding chemotype A, B, and C is that they have several chiral centers implying they are rather difficult for synthesis. Though precision of the ADMET prediction methods is limited, it is possible that some of the molecules in Table 6 which passed all the filters turn to be safe and effective drugs.

REFERENCES

1. Lambert, D.; Hooper, N.; Turner, A. Biochem. Pharmacol. 2008, 75, 781.
2. Guy, J.; Jackson, R.; Acharya, K.; Sturrock, E.; Hooper, N.; Turner, A. Biochemistry 2003, 42, 13185.
3. Burrell, L. M.; Burchill, L.; Dean, R. G.; Griggs, K.; Patel, S. K.; Velkoska, E. Exp. Physiol. 2012, 97(4), 477-485.
4. Byrnes, J.; Gross, S.; Ellard, C.; Connolly, K.; Donahue, S.; Picarella, D. Inflamm. Res. 2009, 58(11), 819-27.
5. Feng, Y.; Gao, G. Comp. Immun. Microbiol. Infect. Dis. 2007, 30(5-6), 309-327.
6. Li, W.; Moore, M.; Vasilieva, N.; Sui, J.; Wong, S.; Berne, M.; Somasundaran, M.; Sullivan, J.; Luzuriaga, K.; Greenough, T.; Choe, H.; Farzan, M. Nature 2003, 426, 450.
7. Hofmann, H.; Pyrc, K.; van der Hoek, L.; Geier, M.; Berkhout, B.; Pohlmann, S. Proc. Natl. Acad. Sci. USA 2005, 102, 7988.
8. Kuba, K.; Imai, Y.; Rao, S.; Gao, H.; Guo, F.; Guan, B.; Huan, Y.; Yang, P.; Zhang, Y.; Deng, W.; Bao, L.; Zhang, B.; Liu, G.; Wang, Z.; Chappell, M.; Liu, Y.; Zheng, D.; Liebbrandt, A.; Wada, T.; Slutsy, A.; Liu, D.; Qin, C.; Jiang, C.; Penninger, J. Nat. Med. 2005, 11, 875.
9. Han, D.; Penn-Nicholson, A.; Cho, M. Virology 2006, 350, 15.
10. Li, W.; Sui, J.; Huang, I.; Kuhn, J.; Radoshitzky, S.; Marasco, W.; Choe, H.; Farzan, M. Virology 2007, 367, 367.
11. Pyrc, K.; Bosch, B.; Berkhout, B.; Jebbink, M.; Dijkman, R.; Rottier, P.; van der Hoek, L. Antimicrob. Agents Chemother. 2006, 50, 2000.
12. Der Sarkissian, S.; Huentelman, M.; Stewart, J.; Katovich, M.; Raizada, M. Prog. Biophys. Mol. Biol. 2006, 91, 163.
13. Towler, P.; Staker, B.; Prasad, S.; Menon, S.; Tang, J.; Parsons, T.; Ryan, D.; Fisher, M.; Williams, D.; Dales, N.; Patane, M.; Pantoliano, M. J. Biol. Chem. 2004, 279, 17996.
14. Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. Nucleic Acids Res. 2000, 28, 235.
15. Dales, N.; Gould, A.; Brown, J.; Calderwood, E.; Guan, B.; Minor, C.; Gavin, J.; Hales, P.; Kaushik, V.; Stewart, M.; Tummino, P.; Vickers, C.; Ocain, T.; Patane, M. J. Am. Chem. Soc. 2002, 124, 11852.
16. Tripos Inc. 1699 South Hanley Rd.: St. Louis, Missouri, 63144, USA.
17. Wang, R.; Lu, Y.; Wang, S. J. Med. Chem. 2003, 46, 2287.
18. Jones, G.; Willett, P.; Glen, R.; Leach, A.; Taylor, R. J. Mol. Biol. 1997, 267, 727.
19. Muegge, I.; Martin, Y. J. Med. Chem. 1999, 42, 791.
20. Kuntz, I.; Blaney, J.; Otley, S.; Langridge, R.; Ferrin, T. J. Mol. Biol. 1982, 161, 269.
21. Eldridge, M.; Murray, C.; Auton, T.; Paolini, G.; Mee, R. J. Comput.-Aided Mol. Des. 1997, 11, 425.
22. Cramer, R.; Patterson, D.; Bunce, J. J. Am. Chem. Soc. 1988, 110, 5959-5967.
23. Kapetanovic, I. Chem. Biol. Interact. 2008, 171, 165.
24. Sander, T. Actelion Pharmaceuticals Ltd. Gewerbestrasse 16, 4123 Allschwil, Switzerland.
25. Lipinski, C. Drug Discovery Today: Technol. 2004, 1, 337.
26. Van de Waterbeemd, H.; Gifford, E. Nat. Rev. Drug Disc. 2003, 2, 192.
27. Fülöp, V.; Böcskei, Z.; Polgár, L. Cell 1998, 94, 161.
28. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S.; Johnson, A. J. Mol. Graph Model. 2006, 7, 421.
29. Guy, L.; Lambert, D.; Warner, F.; Hooper, N.; Turner, A. Biochim. et Biophys. Acta 2005, 1751, 2.
30. Schecther, I.; Berger, A. Biophys. Res. Com. 1967, 27, 157.
31. Asif, M. Curr. Med. Chem. 2012, 19, 2984-2991.
32. Aronov, A. M. Drug Discov. Today 2005, 10, 149.
33. Nair, P.; Sobhia, E. J. Mol. Graph Model. 2007, 26, 117.
34. Dong, X.; Zhang, Z.; Wen, R.; Shen, J.; Shen, X.; Jiang, H. Bioorg. Med. Chem. Lett. 2006, 16(22), 5913-5916.