Pergamon

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry Letters 13 (2003) 3989–3992

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

Identifying Inhibitors of the SARS Coronavirus Proteinase

Ekachai Jenwitheesuk and Ram Samudrala*

Computational Genomics Group, Department of Microbiology, University of Washington School of Medicine, Seattle, WA 98195, USA

Received 18 June 2003; revised 23 August 2003; accepted 27 August 2003

Abstract—The Severe Acute Respiratory Syndrome (SARS) is a serious respiratory illness that has recently been reported in parts of Asia and Canada. In this study, we use molecular dynamics (MD) simulations and docking techniques to screen 29 approved and experimental drugs against the theoretical model of the SARS CoV proteinase as well as the experimental structure of the transmissible gastroenteritis virus (TGEV) proteinase. Our predictions indicate that existing HIV-1 protease inhibitors, L-700,417 for instance, have high binding affinities and may provide good starting points for designing SARS CoV proteinase inhibitors.

A novel coronavirus (CoV) has been isolated and identified as the cause of the severe acute respiratory syndrome (SARS),[1] for which there is currently no effective treatment. The SARS CoV genome sequence has been recently published.[2] The structure of the main proteinase, essential for SARS CoV replication, can be deduced from its similarity (43% sequence identity) to the X-ray crystallography structure of the proteinase from the porcine transmissible gastroenteritis virus (TGEV), also a coronavirus.[3]

Anand et al., who solved the structure of the TGEV proteinase (PDB code: 1lvo),[4] have produced a model for the SARS CoV proteinase (PDB code: 1p9t and 1pa5) using the former as a template. In addition, they propose a drug discovered to treat the common cold, AG7088, as a good starting point for SARS CoV inhibition. AG7088 by itself has failed to inhibit the SARS CoV in vitro.[5] Here, using simulations, we show that HIV-1 protease inhibitors may provide better starting points than AG7088 for inhibition of the SARS CoV proteinase.

Our computational simulation approaches use molecular dynamics (MD) and docking techniques, and have been applied to calculate the binding affinities of HIV-1 protease mutants and their inhibitors.[6] The affinities have then been used to predict resistance/susceptibility for 1800 HIV-1 protease mutants/inhibitor combinations with greater than 90% accuracy. In a similar fashion, we have calculated the binding affinities of a variety of known HIV-1 protease inhibitors to the TGEV main proteinase.

Docking calculations were carried out using AutoDock version 3.0.5, with the Larmarckian genetic algorithm (LGA).[7] Since the structures of the TGEV and SARS CoV proteinases–inhibitor complexes have not been obtained, we therefore first performed preliminary docking experiments to identify the potential binding sites of the inhibitors by generating a grid box that is big enough to cover the entire surface of the protein. Docking runs were set to 100 to allow AutoDock to exhaustively find the potential binding sites of each inhibitor. The first ranked docking solution showed that all inhibitors bound to the substrate binding site of the TGEV and SARS proteinases.

The protein–inhibitor complexes derived from the preliminary docking step were then immersed in TIP3-water shell and all atoms were allowed to relax using MD simulation. MD simulations were carried out with the NAMD software version 2.5b1[8] using the X-PLOR force field. [9] One hundred steps of energy minimization of the protein–inhibitor–water complex were initially performed, followed by 0.1 picoseconds (ps) MD simulation at 300 K, with an atom-based shifted distance-dependent dielectric constant, $\varepsilon = 4r$; a switch function on van der Waals interaction, and a time step of 1 femtosecond (fs). The nonbonded interaction list was

*Corresponding author. Tel.: +1-206-732-6122; fax: +1-206-732-6055; e-mail: ram@compbio.washington.edu

0960-894X/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2003.08.066

updated every 20 time steps. The van der Waals interactions were truncated at a distance of 12 Å. The simulations were repeated with four different starting seeds. The trajectories at 0.1 ps, were recorded and processed in the second docking step.

To calculate the binding affinities of each inhibitor, the 3D affinity grid box was created using the sulphur atom of the cysteine residue of the catalytic dyad (Cys 144 in TGEV proteinase and Cys 145 in SARS proteinase) as a grid center. The number of grid points in the *x*-, *y*-, *z*-axes was 100×100×100 with grid points separated by 0.375 Å. Docking calculations were set to 100 runs. At the end of the calculation, AutoDock performed cluster analysis. Docking solutions with ligand all-atom root mean square deviation (RMSD) within 1.0 Å of each other were clustered together and ranked by the lowest energy representative. The lowest-energy solution was accepted as the calculated binding energy. The inhibitory constant (K_i) was calculated from the final docked energy according to Hess's law and used to define the binding affinity of the inhibitors. More details of our MD and docking protocols are available elsewhere.[6]

Figure 1 lists our predicted binding affinities of 28 known inhibitors of HIV-1 protease as well as the compound proposed by Anand et al., to the TGEV proteinase structure. Of the 29 compounds evaluated, the inhibitor AG7088 is ranked 24th with low predicted binding affinity. On the other hand, several HIV-1 protease inhibitors bind to the substrate binding site of the TGEV proteinase with higher affinity. The structural similarity among these inhibitors can be categorized into two groups: The first group includes inhibitors that resemble the cyclic urea HIV-1 protease inhibitor, that is L-700,417, XK2, LDC, LGZ and DMP. The second group includes those molecules that mimic the peptide backbone conformation, that is, TPV, INU, 1IN, IDV, NFV and LPV. An identical ranking of binding affinities for the inhibitors is obtained if different random starting seeds for the simulations are used.

Figure 2 compares the predicted binding modes of the TGEV proteinase to the inhibitor proposed by Anand et al., (Fig. 2a, in orange) and the inhibitor predicted to have the highest binding affinity (Fig. 2b, in green). The figures show that the AG7088 inhibitor is predicted to partially fit into the binding pocket that the substrate would occupy, whereas the inhibitor with the highest binding affinity, L-700,417 (VAC),[10] mimics the substrate binding mode very well.

L-700,417 is a pseudo C_2 symmetry molecule (Fig. 3), composed of benzyl (P1) and hydroxyaminoindan (P2) groups on each side of the molecule with one hydroxyl group at the center (Fig. 3). AutoDock arranges the hydroxyaminoindan (P2) and the benzyl group (P1′) into S1 and S2 specificity pockets of the TGEV proteinase, respectively, while the P2′ group is inserted into an antiparallel β-sheet that is formed by residues 164–167 and 186–191. Several intermolecular hydrogen bonds are formed between L-700,417 and the enzyme. The carbonyl oxygen (O15 and O36) of the inhibitor form hydrogen bonds to Nε atom of His 41 and to the amide nitrogen of Glu 165, respectively. A water molecule (W681) that is in between the hydroxyl oxygen (O22) of the indanol amide of the inhibitor and the amide nitrogen of Glu 165 forms a hydrogen bridge to the protein as well as the inhibitor. An additional hydrogen bond is formed between the hydroxyl oxygen (O70) of the indanol amide to the amide nitrogen of residue Ser 189.

In contrast to L-700,417, AG7088 (Fig. 3) has poor hydrogen bonding to the substrate binding site. The *p*-fluorobenzyl group is placed in the S1 subsite and the lactone derivative of glutamine is pointed into S2 subsite. The remaining parts of AG7088 'float' above the

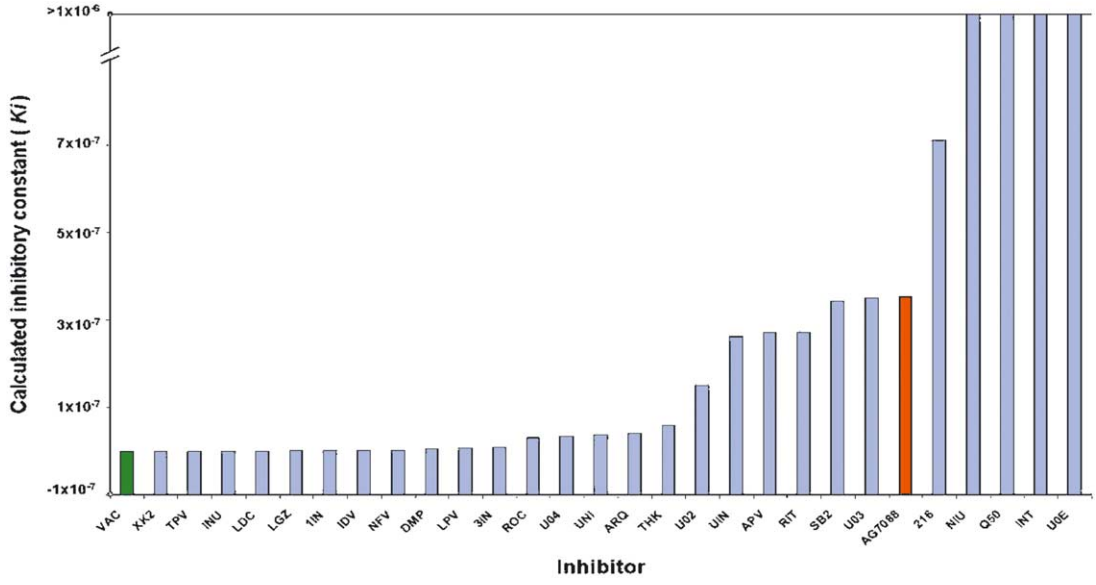

Figure 1. Predicted binding affinities of 29 compounds to the TGEV proteinase. The calculated binding energies are converted into inhibitory constants (K_i) to give a measure of the binding affinities (the lower the K_i the greater the binding affinities). AG7088, shown in orange, is the inhibitor proposed by the authors of the model of the SARS CoV proteinase.[3] The compound with the highest binding affinity is L-700,417, shown in green.

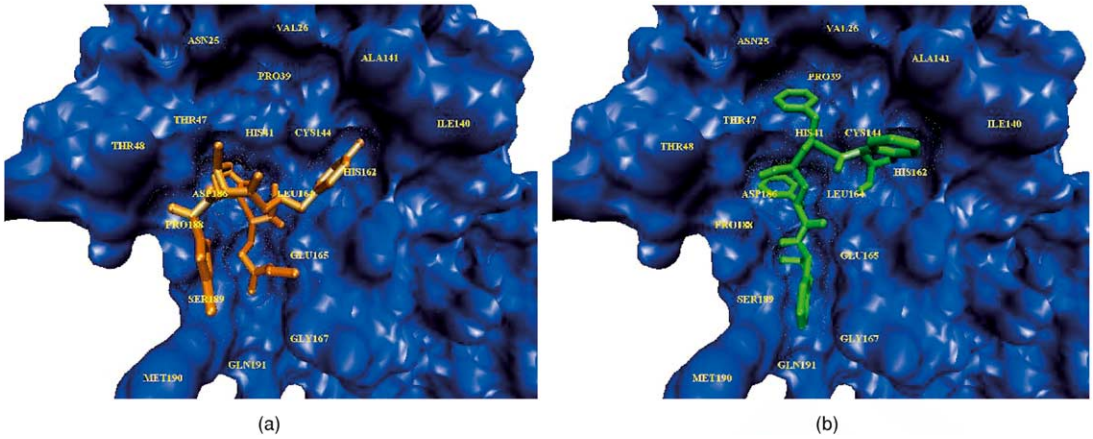

Figure 2. Comparison of the predicted binding modes of the TGEV proteinase and its putative inhibitors: (a) depicts the binding mode of the inhibitor, AG7088, proposed by Anand et al., to the TGEV proteinase structure; (b) depicts the binding mode of the inhibitor L-700,417 (VAC)[10] predicted to have the highest affinity using the molecular dynamics simulation and docking techniques.[6] The AG7088 inhibitor only partially fits into the binding pocket that the substrate would occupy, whereas the inhibitor with the highest binding affinity, L-700,417, mimics the substrate binding mode very well.

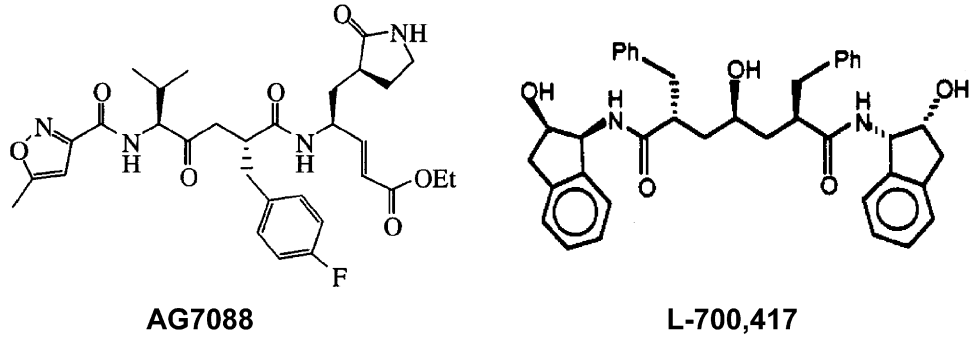

Figure 3. Chemical structures of AG7088 and L-70,417 (VAC).

binding pocket surface with no hydrogen bond formation to the adjacent residues.

We use the TGEV proteinase structure here primarily for illustration purposes, since the substrate binding mode has been elucidated in detail by the crystallographers.[4] However, similar results were also obtained when theoretical models of SARS CoV proteinase (PDB codes 1p9t and 1pa5) were considered.

The experimental observation that AG7088 has been shown to not bind with high affinity to the SARS CoV proteinase[5] is in concordance with our predictions. Further, Anand et al, do not claim that AG7088 as is would work in the case of SARS CoV proteinase (in fact, they only claim that "it will be a very good starting point").[3] This study illustrates that other compounds, for instance L-700,417, may be better starting points to enable discovery of anti-SARS drugs, as well as the utility of computational modelling in evaluating the binding affinities for novel targets and compounds.

Acknowledgements

This work was supported in part by a Searle Scholar Award to Ram Samudrala. We thank members of the Samudrala group for helpful comments.

References and Notes

1. Peiris, J. S.; Lai, S. T.; Poon, L. L.; Guan, Y.; Yam, L. Y.; Lim, W.; Nicholls, J.; Yee, W. K.; Yan, W. W.; Cheung, M. T.; Cheng, V. C.; Chan, K. H.; Tsang, D. N.; Yung, R. W.; Ng, T. K.; Yuen, K. Y. *Lancet* **2003**, *361*, 1319.
2. Marra, M. A.; Jones, S. J.; Astell, C. R.; Holt, R. A.; Brooks-Wilson, A.; Butterfield, Y. S.; Khattra, J.; Asano, J. K.; Barber, S. A.; Chan, S. Y.; Cloutier, A.; Coughlin, S. M.; Freeman, D.; Girn, N.; Griffith, O. L.; Leach, S. R.; Mayo, M.; McDonald, H.; Montgomery, S. B.; Pandoh, P. K.; Petrescu, A. S.; Robertson, A. G.; Schein, J. E.; Siddiqui, A.; Smailus, D. E.; Stott, J. M.; Yang, G. S.; Plummer, F.; Andonov, A.; Artsob, H.; Bastien, N.; Bernard, K.; Booth, T. F.; Bowness, D.; Czub, M.; Drebot, M.; Fernando, L.; Flick, R.; Garbutt, M.; Gray, M.; Grolla, A.; Jones, S.; Feldmann, H.; Meyers, A.; Kabani, A.; Li, Y.; Normand, S.; Stroher, U.; Tipples, G. A.; Tyler, S.; Vogrig, R.; Ward, D.; Watson, B.; Brunham, R. C.; Krajden, M.; Petric, M.; Skowronski, D. M.; Upton, C.; Roper, R. L. *Science* **2003**, *300*, 1399.
3. Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. *Science* **2003**, *300*, 1763.
4. Anand, K.; Palm, G. J.; Mesters, J. R.; Siddell, S. G.; Ziebuhr, J.; Hilgenfeld, R. *EMBO J.* **2002**, *21*, 3213.
5. Clarke, T. *Nature* (Science Update): http://www.nature.com/nsu/030512/030512-11.html.
6. Jenwitheesuk, E.; Samudrala, R. *BMC Struct. Biol.* **2003**, *3*, 2.
7. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.;

Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639.
8. Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283.
9. Brunger, A. T. In *X-PLOR version 3.1, A System for X-ray Crystallography and NMR*. Yale University Press: New Haven, CT 1992.
10. Bone, R.; Vacca, J. P.; Anderson, P. S.; Holloway, M. K. *J. Am. Chem. Soc.* **1991**, *113*, 9382.