Journal of Computational Intelligence in Bioinformatics
ISSN: 0973-385X Volume 3 Number 2 (2010) pp. 229–238

http://www.ripublication.com/jcib.htm

Structure Based Drug Designing Against Putative Targets of Severe Acute Respiratory Syndrome (SARS) Coronavirus and their Computational Evaluation

Vijay Kumar[1], Swati Goswami[1], Rajnikant Namdeo[2] and Reena Jain[*]

*[*1]Department of Microbiology, Boston College for Professional Studies, Gwalior (M.P.), Pin-474006, India*
[2]IBI Biosolutions Private Limited, SCO-7, Swastik Vihar, Mansa Devi Complex-5, Punchkula (Haryana), Pin-134109, India
[*]Corresponding Author E-mail: rina31in@yahoo.co.in

Abstract

This study was aimed to develop promising drug molecules against human SARS coronavirus. The four major proteins, Spike, Helicase, 3CLpro and RdRp that play crucial role in viral life cycle were taken to be the putative targets for drug development through structure based drug designing (SBDD). 3-D structures of target protein were generated by homology modeling using MODELER. "CHEMSKETCH" and "LIGBUILDER" software were used for generating ligand molecules against most active sites of each target. The Ligand molecules having lowest docking energy (LigS-7 -13.11 Kcal/mol; LigR-6 -10.01Kcal/mol; LigH-5 -172354.75Kcal/mol; LigP-3 -410013.45Kca/mol) were validated for their possible use as drug using validation tools MOLINSPIRATION and OSIRIS. OSIRIS showed developed candidate drugs LigS-7, LigR-6, LigH-5 and LigP-3 to be non-allergic, non-mutagenic, non-tumorigenic and safe for reproductive system. Molinspiration results suggested these drugs to be an active molecule with low molecular weight, non allergic, good absorption by tissues, safe to body system and followed Lipinski's rule of five for a potent drug. The IUPAC name of these drugs active against spike protein, Helicase, 3CLpro and RdRp are {[(1*S*,3*R*)-6-cyclopentyl-1,3-dimethylhexyl]amino}acetaldehyde, 2[(hexylamino)methyl]heptanoic acid, 2-{[(4-hexylpyrrolidin-2-yl)methyl]amino}-2-oxoethyl acetate and (2*R*,3*R*,6*S*)-3-amino-2-ethyl-6-methyldecanoic acid respectively.

Keywords: Heptad repeat regions, non structural proteins, SBDD.

Introduction

In the first half of 2003, the global community saw the emergence and impact of severe acute respiratory syndrome (SARS), the first severe and easily transmissible new infectious disease of the new millennium. From Guangdong province of China, SARS virus spread along international travel routes to 30 countries and became deeply embedded in 6 of them. By 11th July 2003, 8437 probable SARS cases were reported from 29 countries with 813 deaths**.** Dr. Carlo Urbani, a WHO epidemiologist who investigated the Hanoi outbreak, was the first to recognize the condition as a distinct entity **[1, 2, 3, 4, 5]**. Severe acute respiratory syndrome or SARS is a respiratory disease in humans which is caused by the SARS coronavirus (SARS CoV). SARS-CoV entry into target cells is initiated by binding of viral spike (S1) glycoprotein with receptors Angeotensin converting enzyme-2 (ACE-2) present on host epithelial cell and fusion with host membrane is mediated by S2 region of spike glycoprotein. The coronavirus spike protein is characterized by the presence of two heptad repeats, HR1 and HR2 which forms a fusion-active hairpin conformation **[6, 7, 8]**. This hairpin structure likely juxtaposes the viral and cellular membranes, thus facilitating membrane fusion and subsequent viral entry **[9, 10]**. Immediately after entry, host RNA polymerase translates the two third gene from 5′ end of viral genomic RNA to produce two polyproteins pp1a and pp1b (replicase complex) which have proteolytic activity**[11]**. These two initial translation products are then auto-proteolytically processed primarily by the main proteinase (3CLpro), encoded within the 5'-proximal region of the replicase gene to release a number of nonstructural proteins including the RNA-dependent RNA polymerase (RdRp) and the Helicase (nsp10). RdRp transcribes the genome to produce five to seven subgenomic mRNAs present in RNA duplex form (+ve and –ve sense RNA duplex). Helicase unwinds the RNA duplex and mRNA translates to produce 16 non functional proteins. 3CLpro cleaves these poly-proteins to generate functional proteins[12, 13].

This study was under taken to develop drug molecules against the four major proteins, spike, helicase, 3CLpro and RdRp that play crucial role in viral life cycle through structure based drug designing (SBDD).

Methodology

Amino acid sequence for all four target proteins of Human SARS CoV were obtained from NCBI. Accession numbers were GI: 30173397, GI:30027621, GI: 37999886, GI: 82408235 for spike, RdRp, helicase and 3CLpro respectively **[14]**. These were taken as query sequence.

Three dimensional structure prediction of target protein

PSI- BLAST was performed by taking the FASTA format of amino acid sequence for each targets protein separately to search for the protein having highest genetic and structural similarity. In the PSI-BLAST result, proteins showing highest similarity to our query sequence were identified and searched for aminoacid sequences in Protein Data Bank (PDB) and one of these was taken as our template sequence. With the help

of query sequence and template sequence, homology modeling was performed using MODELER.

Structure validation

Structures generated from each of our target sequence were evaluated using protein structure analysis and validation tool, PROCHECK. It is a suite of programs to check the stereo chemical quality of protein structures. It gives the details related to bond angle, bond length, dihedrals, covalent bonds and chirality of the molecule and uses Ramachandran's plot to give dihedral angles ϕ against φ of amino acids residues in the protein structure**.**

The best model having lowest number of Energy minimization cycles was selected for each of the four targets and cavities were identified by running PASS **[15].** PASS (Putative Active Sites with Spheres) is a simple computational tool that uses geometry to characterize regions of buried volume in proteins and to identify positions likely to represent binding sites based upon the size, shape, and burial extent of these volumes.

Ligand development

Swiss-Pdb viewer (v4.0.1), software developed by Swiss Institute of Bioinformatics, for visualizing the 3-D structure of protein **[16]**, was used to view the 3-D structure of individual targets and to identify amino acid residue closest to the most active site in each protein. This was exploited for binding of the ligand molecule. Structure of existing inhibitors for the template protein was searched using PDB and the seed molecules were designed in "CHEMSKETCH" against all target proteins. CHEMSKETCH is software developed for 3-D structure development, predicting physical and chemical properties, generating IUPAC nomenclature etc. of a chemical compound **[17].** Software "HEX" was used to dock the seed molecule on the amino acid residue nearest to the most active site**.** Seed molecules were grown into complete ligand molecule within the cavity according to the availability of growing sites using software LIGBUILDER **[18].**

Ligand docking

Developed Ligand molecules were again docked into the cavity using docking tool, "AUTODOCK" and results were viewed and analyzed for lowest docking energy**.** AUTODOCK helps to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure **[19]**.

Drug validation

Ligand molecule with lowest docking energy was validated using two online validation tools, MOLINSPIRATION and OSIRIS Property Explorer for evaluation

of biological properties [20, 21, 22]. MOLINSPIRATION calculates the biological activities of a drug molecule with respect to following: LogP, TPSA, GPCR, molecular weight, violations, ion channel modulator, kinase inhibitor. OSIRIS property explorer deals with toxicity risk (like mutagenic, tumorigenic, irritant, effect on reproductive system) LogP, solubility, molecular weight, druglikeness and drug score.

Results

The proteins (pdb|2GHV|E, pdb|2H85|A, pdb|2FYG|A and pdb|2C3S) had more than 95% similarity with our query sequence of Spike protein, RdRp, Helicase and 3CLpro respectively in PSI-BLAST. 3-D models of target protein generated by homology modeling and cavities present on the protein surface are shown in **figure 1**. Results of homology modeling, PASS, LIGBUILDER and docking energy of generated ligand molecules calculated by AUTODOCK for each target protein are summarized in **Table 1**.

Biological behavior of ligands for possible use as drug molecules by two online validation tools MOLINSPIRATION and OSIRIS property calculator is summarized in **Table 2**. The ligands; LigS-7, LigR-6, LigH-5 and LigP-3 active against Spike protein, RdRp, Helicase and 3CLpro respectively were found to be non-mutagenic, non-tumorigenic and safe for reproductive system. Molinspiration results suggested these drugs to be an active molecule with low molecular weight, good absorption by tissues, safe to body systems and followed Lipinski's rule of five for a potent drug. Out of these, except drug molecules LigR-6 active against RdRp, all other drug molecules were shown to be allergic for the biological system and had low druglikeness and drug score values.

To improve biological properties, from the drug molecule LigS-7 alcoholic group (-OH) at 3rd carbon position was removed. In LigH-5, a methyl group was introduced at C1. While for LigP-3, a methyl group from C5 and the Oxo (=O) group from C7 was removed. The IUPAC name and structure of the modified ligand molecules generated against each targets, are given in **Table 3**. Docking energies and biological properties of these modified drug molecules are summarized in **Table 4.** Modifications removed irritant property and increased druglikelyness.

Discussion

Developing a structurally compliment molecule with appropriate biological activity against target molecule was the main aim of this study. Goal of attaining a molecule with suitable biological activity was kept under the strict observation by considering the Christopher Lipinski's rule-of-five as the prime criteria during the whole procedure. Among the developed molecule for all four target molecules (Spike, RdRp, Helicase and 3CLpro); molecule LigS-7, LigR-6, LigH-5 and LigP-3 got maximum drug score (0.48, 0.46 0.46 and 0.40 respectively) over in silico validation. Drug score suggests the overall fitness among biological parameters such as solubility or a good intestinal absorption for an oral drug. The accountability of 2-D drug likeness

(Structure Activity Relationship) was also considered for drug score calculation. LigS-7, LigR-6, LigP-3 and LigH-5 were selected as best on the basis of their interaction with their target molecule. The interaction is quantified in terms of Gibbs energy of a reaction between the target molecule and the inhibitor molecule. For a reaction being spontaneous it must be negative. The interaction energy, calculated for LigS-7, LigR-6, LigP-3 and LigH-5 was found minimum

(-13.11, -10.01, -410013.45 and -172354.75 Kcal/mol respectively) among other developed molecules. This suggests that how intense the interaction will be. Interaction energy depends upon the chemical competency between the two interacting molecules. Online validation of biological properties such as a mutagenic nature, carcinogenicity, effect on reproductive system and an irritant nature were also measured for all the developed ligand molecules. All ligand molecules, LigS-7, LigR-6, LigP-3 and LigH-5 were found safe for human use; however, LigS-7, LigP-3 and LigH-5 molecules pursue irritant nature. Mutations were done in accordance to remove this irritant effect from molecule LigS-7, LigP-3 and LigH-5 and it was found that removal of "-OH" group improves the properties of LigS-7 by decreasing the irritation effect (lowering allergic responses to acceptable range), thereby increasing the drug likeness value to 0.13 and drug score to 0.48. Similarly removal of a methyl group from C5 and the Oxo (=O) group from C7 of LigP-3 molecule and addition of methyl group to LigH-5 also decreased their irritant to acceptable range with drug scores 0.46 and 0.45 respectively. Occurrence of this irritant nature in molecule indicates that it may because of inappropriate chemical nature of lead molecule selected for the development of inhibitors. It is suggested that in future development consideration of same lead molecule may produce the same effect of irritation in developing molecule. A more refined lead molecule is a prime requirement for the future workers using the same approach.

Possible side effects of these probable drug molecules were calculated. Major groups of biological molecules taken to study the side effects included GPCR ligands, Ion channel modulators etc. Calculation of these values for all designed ligand molecules suggested that these molecules had no tendency to work as GPCR ligand, Ion channel modulator, Kinase inhibitor or as a nuclear receptor ligand. LogP (Moriguchi's correction) was evaluated to ensure the absorption and availability of possible drug molecule. It was found as 3.86 for LigS-7, 4.01 for LigP-3, 2.52 for LigR-6 and 2.12 for LigH-5 for the corrected molecules using mutations. These results suggest the values are according to "rule of 5". Optimum value of logP for a drug is accepted between 2.5 to 4. Solubility was also measured and found better for all four molecules. Molecular weight were found under the limit, as defined by "Rule of 5" (<500, Table-4). All this ensures the good absorption of these molecules at the intestinal surface or transport across the cell membrane.

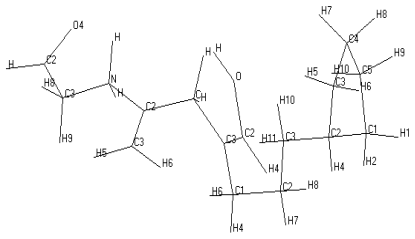
Above findings for developed molecules LigS-7, LigR-6, LigP-3 and LigH-5 with IUPAC name {[(1*S*,3*R*)-6-cyclopentyl-1,3-dimethylhexyl]amino}acetaldehyde, (2*R*,3*R*,6*S*)-3-amino-2-ethyl-6-methyldecanoic acid, 2-[(hexylamino)methyl]heptanoic acid, 2-{[(4-hexylpyrrolidin-2-yl)methyl]amino}-2-oxoethyl acetate respectively suggest them to be used as good drug molecules to treat the infection of SARS virus in future. Such Structure-based drug design of protein

ligands has emerged as a new tool in medicinal chemistry [23]. *In silico* methods have the potential to reduce both time and cost in developing suggestions on drug/ lead-like molecules. The greatest success of computer-aided structure-based drug design to date is the HIV-1 protease inhibitors that have been approved by the United States Food and Drug Administration and reached the market [24]. This itself highlights the importance of the study.

Table 1: Results of Homology Modeling, Structure validation, PASS and ligand generation.

Target protein	No. of structures generated by Homology modeling	Best model in PROCHECK	Results of PASS		No. of ligand molecules generated by LIGBUILDER	Ligand molecule with lowest docking energy	Docking energy of selected ligand molecule (in kcal/mol)
			Amino acid nearest to the active site	Distance from the active site in Å			
Spike protein	Three (Spk-1, Spk-2, Spk-3)	Spk-2	Asn 108	3.06	Seven (LigS-1, to LigS-7)	LigS-7	-13.11
Helicases (nsp10)	Five (Hel-1, Hel-2, Hel-3, Hel-4, Hel-5)	Hel-3	Asp 87	3.74	Five (LigH-1, to LigH-5)	LigH-5	-172354.75
Proteinases (3CLpro/Mpro)	Five (Pr-1, Pr-2, Pr-3,Pr-4, Pr-5)	Pr-4	Asn 151	3.46	Ten (LigP-1 to LigP-10)	LigP-3	-410013.45
RdRp	Five (Rp-1, Rp-2, Rp-3, Rp-4, Rp-5)	Rp-1	Tyr 279	4.18	Ten (LigR-1 to LigR-10)	LigR-6	-10.01

Table 3: Structure and IUPAC nomenclature of generated ligand molecules for each targets

S.No	Target Protein	Ligand name	Structure	IUPAC nomenclature
1	Spike Protein	LigS-7		{[(1*R*,3*R*)-6-cyclopentyl-3-(hydroxymethyl)-1-methylhexyl]amino}acetaldehyde

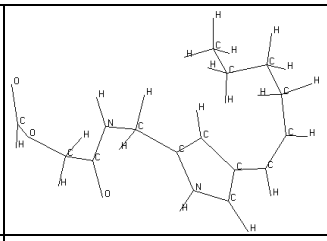
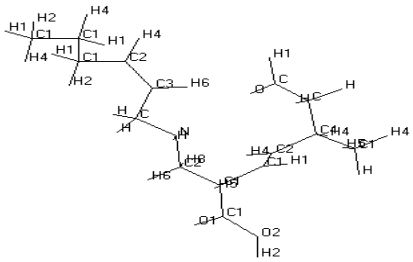
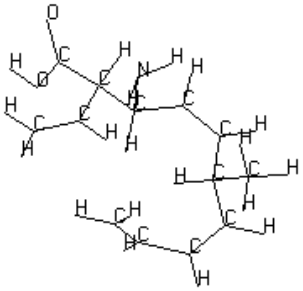
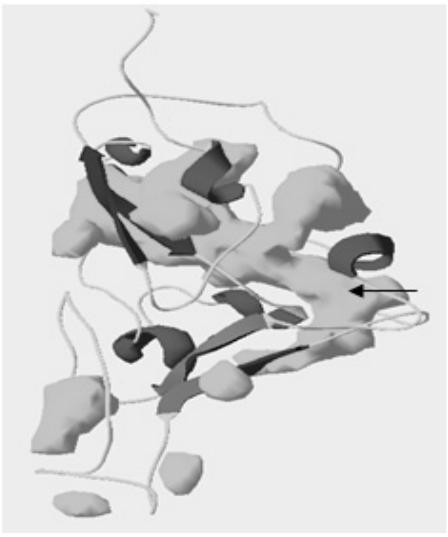
3	Helicase	LigH-5		2-{[(4-hexylpyrrolidin-2-yl)methyl]amino}-2-oxoethyl formate
4	3CLpro	LigP-3		2-[(hexylamino)methyl]-5-methyl-7-oxoheptanoic acid
2	RdRp	LigR-6		(2*R*,3*R*,6*S*)-3-amino-2-ethyl-6-methyldecanoic acid

Table 2: Results of drug validation by online tools MOLINSPERATION and OSIRIS of Original ligand molecules.

Target protein and Ligand molecule	**Tool & Parameter**							
	MOLINSPERATION			**OSIRIS**				
	TPSA	**GPCR**	**Mol. wt**	**Log P**	**Solubility**	**Druglikness**	**Mutagenic**	**Irritant**
Spike protein LigS-7	53.904	-0.11	256.41	2.91	-2.91	-2.88	No	Yes
Helicases LigH-5	67.43	-0.06	270.337	1.84	-2.58	-11.20	No	Yes
Proteinase (3CLpro) LigP-3	66.397	-0.23	271.401	3.4	-2.70	-9.34	No	Yes
RdRp LigR-6	63.322	0.09	229.364	2.52	-2.46	-5.66	No	No

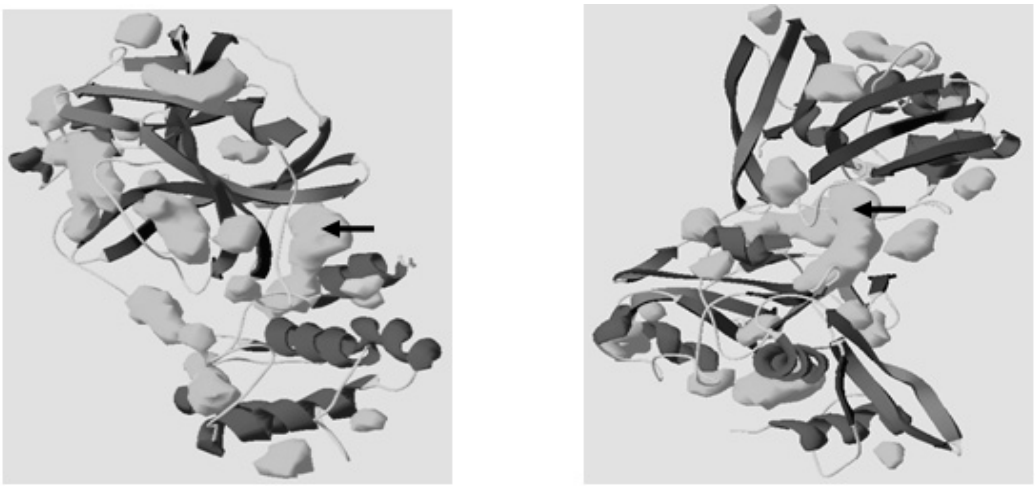
Table 4: Results of drug validation by online tools MOLINSPERATION and OSIRIS of modified ligand molecules.

Target protein and Ligand molecule	**Tool & Parameter**							
	MOLINSPERATION			**OSIRIS**				
	TPSA	**GPCR**	**Mol. wt**	**Log P**	**Solubility**	**Druglikness**	**Mutagenic**	**Irritant**
Spike protein LigS-7	29.098	-0.07	239.40	3.86	-3.42	0.13	No	Within range
Helicases LigH-5	67.430	+0.13	284.40	1.84	-2.58	-12.6	No	No
Proteinase (3CLpro) LigP-3	49.326	+0.20	243.30	4.01	-2.83	-9.34	No	No
RdRp LigR-6	63.322	0.09	229.364	2.52	-2.46	-5.66	No	No



(a) Spk-2

(b) Hel-3



(c) Prt-4 (d) Rp-1

Figure 1: 3-D models with cavities on the surface of each target Proteins as generated by Modeller and viewed under SPDV Spike protein (b) Helicase (c) 3CLpro and (d) RdRp.

Note: In above models, Red is indicating α Helices; Blue β pleated sheats; Light blue coils; Green cavities present on the molecular surface of the protein; Arrows are indicating the position of the most active site sequestered within the surface.

Reference

[1] Drosten, C., Günther, S., Preiser, W., Brodt, H.R., Becker, S., Rabenau, H., et al., 2003, "Identification of a novel coronavirus in patients with severe," The New England journal of medicine, 348(20), pp. 1967-1976.

[2] Guan, Y., Zheng, B. J., He, Y. Q., Liu, X .L., Zhuang, Z .X., Cheung, C. L., et al., 2003, "Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China," Science, 302(5643), pp. 276 - 278.

[3] Ksiazek, T.G., Erdman, D., Goldsmith, C. S., Zaki, S .R., Peret, T., Urbani, C., et al., 2003, "A novel coronavirus associated with severe acute respiratory syndrome," The New England journal of medicine, 348(20), pp. 1953–1966.

[4] Peiris, J. S. M., Lai, S. T., Poon, L. L. M., et al., 2003, "Coronavirus as a possible cause of severe acute respiratory syndrome," The Lancet, 361, pp. 1319–1325.

[5] Severe acute respiratory syndrome (SARS), 2003, "status of the outbreak and lessons for the immediate future," Geneva, World Health Organization.

[6] Jacqueline, D. R., Graham, S., Dhaval, N. G., Andrew, J. R., Scott, L. D., and Paul, B., 2005, "Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry," PNAS, 102 (33), pp. 11876-11881.

[7] Shui, T., Chun, H. H., Tzu, P .K., Hui, M. Y., Twsen, K. T., and Andrew, H. J. W., 2004, "Structural and Immunological characterization of the fusion core of SARS corona virus spike protein," Biochemical and Biophysical Research Communications, 324 (2), pp. 761-767.

[8] Xu, Y., Lou, Z., Liu, Y., Pang, H., Tien, P., George, F. G., et al., 2004, "Crystal Structure of Sever Acute Respiratory Syndrome Coronavirus spike protein fusion core," J. Biol. Chem., 279 (47), pp. 49414-49419.

[9] Shuwen, L., Xiao, G., Chen, Y., He, Y., Niu, J., and Escalante, C., 2003, "Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors," The Lancet, 363 (9413), pp. 938-947.

[10] Supekar, V. M., Bruckmann, C., Ingallinella, P., Bianchi, E., Pessi, A., and Carfí, A**.,** 2004, "Structure of a proteolytically resistant core form the severe acute respiratory syndrome coronavirus S2 fusion protein," PNAS, 101 (52), pp. 17958-17963.

[11] Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., et al., 2005, "Design of wide-spectrum inhibitors targeting coronavirus main proteases," P. LoS Biology., 3(10): e324.

[12] Xiang, X., Yunqing, L., Susan, W., Eddy, A., Stefan, G. S., and Jianping, D., 2003, "Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design," Nucleic Acids Res., 31(24), pp. 7117–7130.

[13] Julian, A. T., Rory, M. W., Yu, B. C., Lin, Y. L., Marie, C. L., and Malik, J. S., 2003, "The SARS Coronavirus NTPase/Helicase Belongs to a Distinct Class of 5' to 3' Viral Helicases," JBC Papers in Press., Published on August 13, (online journal)

[14] http://www.ncbi.nlm.nih.gov/sites/entrez

[15] http://www.ccl.net/cca/software/UNIX/pass/overview.shtml

[16] http://expasy.org/spdbv/bncg

[17] http://www.acdlabs.com/products/chem_dsn_lab/chemsketch/

[18] http://mdl.ipc.pku.edu.cn/drug_design/work/ligbuilder.html

[19] http://autodock.scripps.edu/

[20] http://www.molinspiration.com/

[21] http://www.molinspiration.com/cgi-bin/properties

[22] http://www.organic-chemistry.org/prog/peo/

[23] Klebe, G., 2000, "Recent developments in structure-based drug design," J Mol Med., 78(5), pp. 269-281.

[24] Wlodawer, A., and Vondrasek, J., 1998, "Inhibitors of HIV-1 protease: a major success of structure-assisted drug design," Annu. Rev. Biophys. Biomol. Struct., 27, pp. 249-84.