

GENE (HIV-1 PROTEASE) BASED DRUG (INHIBITOR) DISCOVERY

Swati V. Ghuge, Mukta C. Zarkar, S. R. Deshmukh,

Sanjay K. Choubey

MGM'S College of CS & IT, Department of Biotechnology &
Bioinformatics Nanded (MS) – 431605

Yeshwant College of IT, bioinformatics & Biotechnology Parbhani (MS) – 431401

ABSTRACT:

We can design a drug (inhibitor) to inhibit the activity of gene. As a gene produces protein/enzyme, so to avoid the formation of any disease causing proteins, we have to stop the activity of that gene. With the help of different bioinformatics tools and software's we can do this.

A protease is an enzyme that cleaves proteins to their component peptides. The HIV-1 Protease (PR) hydrolyses viral polyproteins into functional protein products that are essential for viral assembly and subsequent activity. HIV-1 protease activity is critical for the terminal maturation of infectious virions. Once HIV enters the cell, viral RNA undergoes reverse transcription to produce double-stranded DNA (a step inhibited by nucleoside analogues such as zidovudine, didanosine, zalcitabine, stavudine, and lamivudine).

In the presence of HIV-1 protease inhibitors, the virion is unable to mature and is rapidly cleared by poorly understood mechanisms. Figure 1, left, is a photomicrograph of normal budding virions from an infected cell, while Figure 1, right, demonstrates the effect of bathing these cells with the protease inhibitor, saquinavir. The subsequent lack of a dense core for these "ghosted" particles is characteristic of noninfectious HIV virions. By using ncbi we can find the nucleotide and protein sequence of HIV1-Protease. By tool and softwares like pfam, clustalw, gold, blast, we designed the inhibitor "SKF 108738" for HIV1-protease.

Keywords: Inhibitor, HIV1-protease, zidovudine, didanosine, zalcitabine, stavudine, virions, polyproteins.

Introduction:

HIV-1 protease (HIV PR) is an aspartic protease that is essential for the life-cycle of HIV, the retrovirus that causes AIDS. [1][2] HIV PR cleaves newly synthesized polyproteins at the appropriate places to create the mature protein components of an infectious HIV virion. Without effective HIV PR, HIV virions remain uninfected. [3][4] Thus, mutation of HIV PR's active site or inhibition of its activity disrupts HIV's ability to replicate and infect additional

cells, [5] making HIV PR inhibition the subject of much pharmaceutical research. [6]

STRUCTURAL FEATURES:

HIV-1 Protease is a homodimer (chain A, chain B). Each monomer contains 99 amino acids and is identical in conformation. The position of each monomer in the active protease forms an axis of symmetry. The secondary structure of each monomer includes, one

alpha –helix and two anti-parallel beta sheets.

The two Asp25 residues (one from each chain) act as the catalytic residues. According to the mechanism for HIV PR protein cleavage proposed by Jaskolski et al., water acts as a nucleophile, which acts in simultaneous conjunction with a well-placed aspartic acid to hydrolyze the scissile peptide bond.^[7] Additionally, HIV PR has two molecular “flaps” which move a distance of up to 7 Å when the enzyme becomes associated with a substrate.^[8]

Aliphatic residues stabilize each monomer in a hydrophobic core. Additionally, the dimer is stabilized by no covalent interactions, hydrophobic packing of side chains and interactions involving the catalytic residues. Each monomer contains two cysteine residues, but these do not form disulfide bonds.

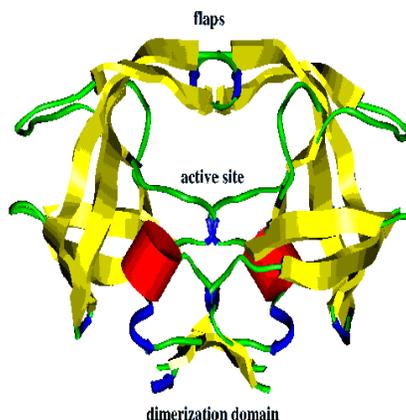
The active site forms at the dimer interface. It is created in a cleft between the two domains as part of four stranded beta turn. The alternate view demonstrates the position of the active site nestled in approximately in the center of the molecule.

An extended turn, a beta hairpin loop, of a beta sheet covers the active site. This flap remains flexible and allows for hinge like mobility. It allows substrate access to the active site by opening and folding the tips into hydrophobic pockets thus exerting a central role in protease activity.

The overall shape of protease is oblong and relatively flat. This surface contour illustrates where potential binding or protein interaction might occur: several binding pockets exists inside the hollow cleft. The structure at left is a ligand.

MECHANISM OF ACTION:-

Fig 1: HIV-1 Protease



HIV-1 protease activity is critical for the terminal maturation of infectious virions. Once HIV enters the cell, viral RNA undergoes reverse transcription to produce double-stranded DNA (a step inhibited by nucleoside analogues such as zidovudine, didanosine, zalcitabine, stavudine, and lamivudine). This viral DNA is integrated into the host genome and, eventually, transcribed and translated by cellular enzymes to produce large, nonfunctional polypeptide chains, referred to as poly-proteins. After these poly-proteins are assembled and packaged at the cell surface, immature virions are produced and released into the plasma. At this point, HIV-1 protease, acting as a "molecular scissors," cleaves the poly-proteins into smaller, functional proteins, thereby allowing the virion to mature. In the presence of HIV-1 protease inhibitors, the virion is unable to mature and is rapidly cleared by poorly understood mechanisms. Figure 1, left, is a photomicrograph of normal budding virions from an infected cell, while Figure 1, right, demonstrates the effect of bathing these cells with the protease inhibitor, saquinavir. The subsequent lack of a dense core for these

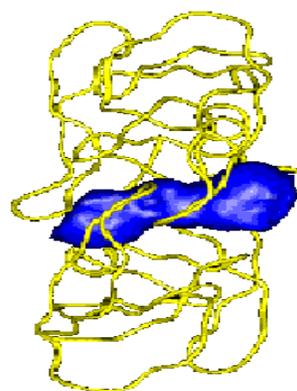
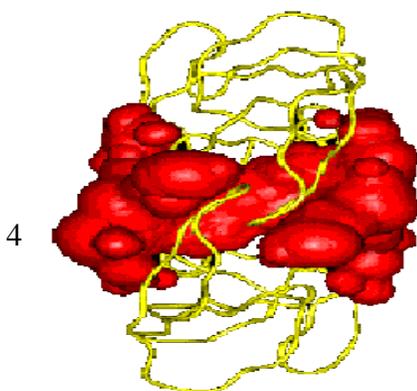
"ghosted" particles is characteristic of noninfectious HIV virions.

[HIV-1 PROTEASE ACTIVE SITE BINDING:-

HIV-1 protease consists of two protein chains. The chains are identical to one another, and each contains 99 amino acids. When the two chains assemble, a long tunnel is formed (as seen from the side view on above left). Protein "flaps" cover the tunnel and open up to allow the enzyme to attach to a protein chain. After attachment, the flaps then close around the protein chain, thereby holding it in the tunnel and allowing the chain to be degraded.

Cheminformatics is the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and optimization.

Since then, both spellings have been used, and some have evolved to be established as Cheminformatics, while European Academia settled in 2006 for Chemoinformatics. The recent establishment of the Journal of Cheminformatics is a strong push towards the shorter variant.



Maximum Volume of Active Site

Minimum Volume of Active Site

fig2:HIV1protease

The illustration in the center shows the flaps in an open conformation and an inhibitor bound to the active site. This is similar to how a protein chain would be bound during protease activity. When the inhibitor is removed as in the illustration on the right, two aspartate residues can be seen (denoted with asterisks), which attack the protein chain and do all the work.

Cheminformatics combines the scientific working fields of chemistry and computer science for example in the area of chemical graph theory and mining the chemical space.^[11, 12] Cheminformatics can also be applied to data analysis for various industries like paper and pulp, dyes and such allied industries.

CHEMINFORMATICS

The term Chemoinformatics was defined by F.K. Brown^[9, 10] in 1998:

COMPUTER AIDED DRUG DISCOVERY

Artificial Intelligence – based drug design supporting systems are employed in this study to identify and in analyzing the

structure activity data. The understanding of the specificity of the biological function is based on the principles of molecular recognition. The binding and action of a drug are controlled by the patterns of molecular fields found in the vicinity of the contact surface of the receptor.

Computational methods employed in this study are on the following patterns.

1. Structure Based Drug Designing.
2. Rational (Analogue) Based Drug Designing.

STRUCTURE BASED DRUG DESIGNING:-

A technique in which 3 dimensional structure of the disease causing molecule is used to design drugs that specifically inactivates its function. Structure Based Drug Design is based on a firm understanding of molecular recognition between active site groups and interacting molecules and is a strategy that has become an integral part of modern drug discovery.

Latest advances in Structure Based Drug Design methodologies including flexible, faster docking techniques, virtual screening and library design. In Structure Based Drug Design, the 3D structure of a drug target interacting with small molecules is used to guide drug discovery. "Structure Based Drug Design represents the idea that how molecule interacts with its target protein."

Structure Based Drug Design can help lead to better compounds more quickly. Structure Based Drug Design is a process whereby the 3D structure of the active site of a biomolecule associated with a disease, as well as the relationship between small molecular structures and their bioactivities are studied. To design a drug against a

specific molecule, the exact structure of that molecule must first be found.

Using computers we can scan through the PDB and find compounds, which can dock with the active sites and thus inhibit the activities of these molecules. Certain computer programs can even design molecules, which fit directly into the active sites of the molecules such as enzymes or receptors, thus making very effective inhibitors.

Materials and Methodology:

GENOMICS

- HIV-1 Protease nucleotide sequence

```
caagggcaaatggtacatcaggccatcacctagaacttt
aatgcaagggtaaaagtagtagaagagaaggctttcag
cccagaagtataccatgtttcagcattatcagaaggag
ccaccacaagatttaaacaccatgctaacacagtggg
gggacatcaagcagccatgcaaatgttaaagagacat
caatgaggaagctgcagaatgggatagagtgcacccag
tgcacaggcctattgcaccaggccagatgagagaacc
aaggggaagtacatagcaggaactactagtaccctcag
gaacaaataggatggatgacaaataatccacatcccag
taggagaaattataaaagatggataatcctgggattaata
aaatagtaagaatgtatagccctaccagcattctggacata
agacaaggacaaaagaacccttagagactatgtagacc
ggttctataaaacttaagagccgagcaagcttcacagga
ggtaaaaaattggatgacagaaacctgttggtcctaaat
cgaaccagattgaagactattttaaagcattgggacca
ggagctacactagaagaatgatgacagcaagct
```

PROTEOMICS

- HIV-1 Protease Protein Sequence
PQITLWQRPLVTIKIGGQLKEALLDTG
ADDTVLEENSLPGRWKPKMIGGIGGFI
KVRQYDQILIEICGHKAIGTVLVGPTP
VNIIGRNLLTQIGCTLNFAAFXVFX

ClustalW:- Multiple Sequence alignment of HIV-1 Protease (www.ebi.ac.uk/tools/clustalw2/index.html)

Pfam:

Pfam is a large collection of multiple alignments – takes HIV-1 Protease protein sequences from Swissprot database, by using SRS retrieval tool and searches for multiple sequence alignment, against the pfam databases.

```
PQITLWQRPLVTIKIGGQLKEALLDTG
ADDTVLEENSLPGRWPKMIGGIGGF
IKVRQYDQILIEICGHKAIGTVLVGPT
PVNIIGRNLLTQIGCTLNFA
AFXFVVX
```

CHEMINFORMATICS

Similarity structure or sub structure searches:- Searches the similarity structure or sub structure of HIV-1 Protease inhibitor (SKF 108738) by using NCI-3D search database.

Draw the HIV-1 Protease Inhibitor:-

We have build the HIV-1 Protease inhibitor 3d structure by using Hyperchem model builder.

Computational methods for HIV-1 Protease Inhibitors (SKF 108738):-

We are using the hyperchem molecular modeling package for computational methods of SKF 108738

Total energy of molecule:-

Submit the 3D structure of SKF 108738 to Hyperchem tool, and then set up the molecular mechanics force field on Hyperchem set up menu. Then click the single point calculations on compute menu of Hyperchem.

Energy minimizations or Geometry optimizations:-

We have minimized the energy of SKF 108738 through by adjusting the geometry and set up the RMS Gradient at 0.1 in Hyper chem Package. Then click the Geometry Optimization in compute menu of hyper chem.

Molecular Dynamics Simulations:-

The solvated molecules of SKF 108738 by using periodic boundary conditions in hyper chem. Then apply the molecular dynamics in compute menu of hyper chem.

- Set up the system
- Starting temperature 100k
- Resulting temperature 300k
- Time step 30k

Monte Carlo Simulations:-

- Applied to the Monte Carlo Simulations.
- Starting temperature 100k
- Resulting temperature 300k
- Time step 30k
- Conformational Analysis:-

We have done the different conformationals structures of SKF 108738 by using Hyperchem. Molecular modeling package.

Activation energy or Transition state structure:-

We have calculated the transition state of SKF 108738 by using Hyperchem. Molecular modeling package.

QSAR Studies:-

We have calculated all the QSAR properties of SKF 108738 by using Hyperchem. Molecular modeling packages.

STRUCTURE BASED DRUG DESIGN:

Structure based drug design is dependent on the existence of a model of the receptor. Many drug receptor interactions are controlled by a few key receptor groups. The goal of SBDD is to try and join these groups together to generate conformationally sensible, synthetically target molecules.

SBDD experimental part is carried out by Ligand Docking (Gold).

Taken some chemical compounds (sk-108738, Paraaminobenzoate, Iodopyrazole, Guanine) from GenomeNet Database Service (www.genome.jp/ligand) by using reference from journal of medicinal chemistry and protein ligand database (<http://www.michell.ch.cam.ac.uk/pld/index.htm>)

Any unusual bonds (disulfide bridges etc.) should have CONECT records.

If a metal ion is present, ensure that all bonds between the ion and co-coordinating protein or water molecules are deleted (GOLD will find them automatically). Metals should be within bonding distance of at least two proteins and /or water molecules in the active site so that GOLD can infer likely coordination geometries.

GOLD: For Docking Set up the GOLD Parameters:-

General settings

- External energy weight : 1.38
- Flood fill probe radius : 1

- Flood fill point spacing : 0.5
- Length of H bond : 2.9
- Maximum Number of operations : 100000
- Doing simplex minimization after every GA run
- Angles coded in 8 bits
- Using ionization dispersion time externally
- Flattening input amide and trigonal nitrogen's
- Solvated point distance : -0.200
- Number of islands : 5
- Population size : 100
- Selection pressure : 1.1
- Flood fill radius : 10
- Flood fill origin
- Cross weight (one string) : 95
- Mutate wt (one string) : 95
- Migration weight : 10
- Niche size : 2

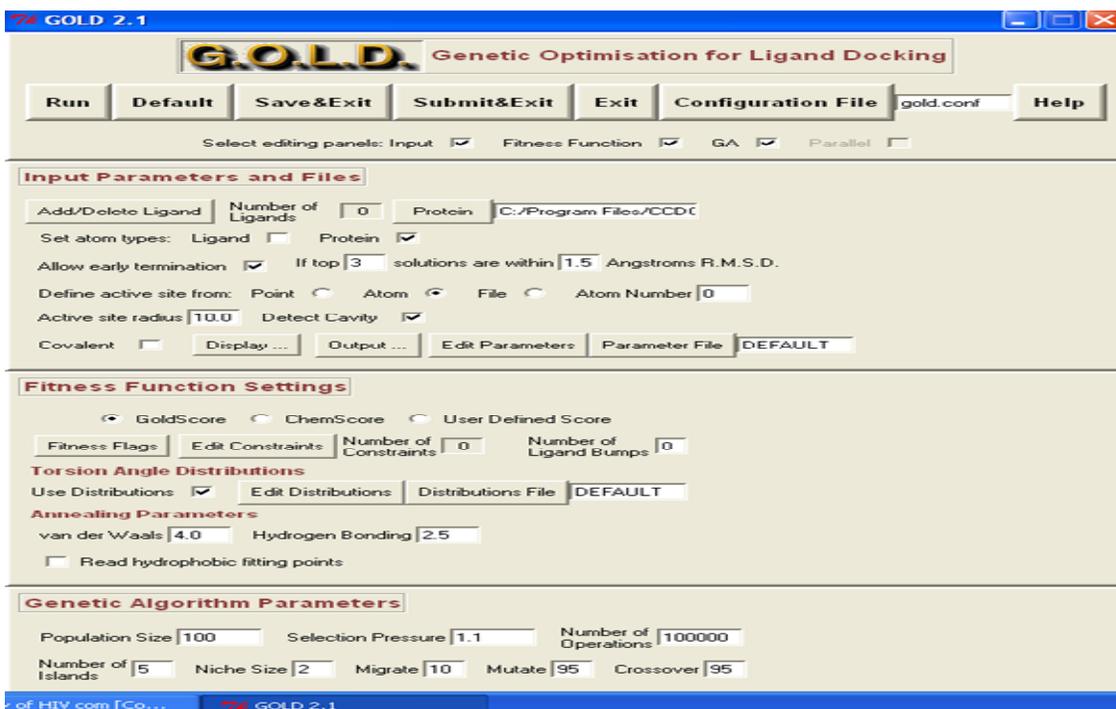


Fig3: Gold homepage

Result: MULTIPLE SEQUENCE ALIGNMENT of HIV 1 PROTEASE :-

>VPRT_SMRVH/172-273
 ISQQRPTLKLKLNK.....LFSGILDT
 GADATVISYTHWPRNWPL..TTVATHL
 RGIGQAT.NPQQAQMLKWE.....DSEG
 NNGHI.....TPYVLPN.LPVN.....LWGRD
 ILSQMKLVMC
 >O41289/167-268
 ITNQKPSLTLWLDGK.....AFTGLID
 TGADVTIKQEDWPSHWPT..TETLTHL
 RGIGQSS.NPKQSSKYLTWT.....DKENN
CHEMINFORMATICS

STRUCTURE OF SKF 108738

SGLI.....KPFVIPY.LPVN.....LWGRDLL
 AQMKIIMC
 >VPRT_JSRV/178-279
 VTEARPELELRINAN.....FFRGVLD
 TGADISVISDKYWPTTWPK..QMAISTL
 QGIGQTT.NPEQSSLLTWK.....DKDGH
 TGQF.....KPYILPY.LPVN.....LWGRDIL
 SKMGVYLY
 >GAG_IPMA/674-774
 SLNDRPKLRLKINGK.....EFEGILDT
 GADKSIISTHWWPKAWPT..TESSHSLQ
 GLGYQS.CPTISSVALTWE.....SSEGQQ
 GKF.....IPYVLP..LPVN.....LWGRDIMQ
 HLGILS

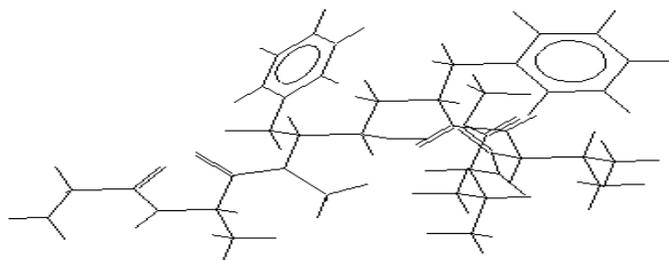


Fig4: 3D STRUCTURE OF SKF 108738

TOTAL ENERGY OF SKF 108738 BY MOLECULAR MECHANICS:-

Total Energy = 543877.312500 Gradient = 758719.187500

Molecule Properties:-

RMS Gradient = $7.587e + 005$ k/cal/Amol

(Gradient x = 536921.4000 k/cal/Amol, Gradient y = 6135162.5000k/cal/Amol,

Gradient z = 1013716.000k/cal/Amol)

Dipole Moment

Total dipole = 88.85417 Dipole x =87.13437, Dipole y = -11.45754, Dipole z = -13.09165

ENERGY MINIMIZATION OF SKF 108738:-

Energy = 83.005363

Gradient = 0.095556, Converged = yes, Cycles = 962, Points = 2035

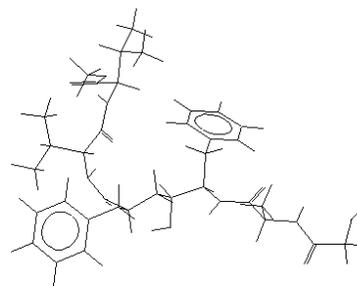
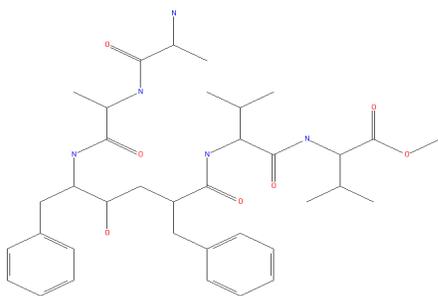


Fig5:ENERGY MINIMISED MODEL

MOLECULAR DYNAMICS SIMULATIONS:-

Time = 1.6ps
 Total Energy = 88.38042 Kcal/mol, T = 283.467K

At t1=0

Ekin = 29.2076Kcal/mol, Epot = -66.43Kcal/mol
 EToT = -52.5204 Kcal/mol, T = 99.98592K

At t2 = 1.6

Ekin = 13.65412Kcal/mol, Epot = -83.00537Kcal/mol
 ETOT = -53.79776Kcal/mol, T = 46.74193K

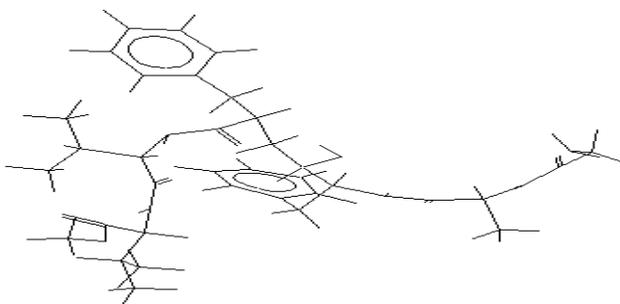


Fig6: MOLECULAR DYNAMICS MODEL

QSAR STUDIES:-

Surface Area (Appr) = 928.77A
 Surface Area (Grid) = 1000.19 A
 Volume of the Molecule = 1874.72 A
 Log P of the molecule = 2.35
 Hydration Energy = 13.52 Kcal/mol
 Refractivity of Molecule = 171.18A
 Polarisability of Molecule = 70.76 A
 Mass of the Molecule = 653.82 amu
 Active site Analysis SKF 108738:-

Solvent accessible cavity atoms:

1204 1277 1558 1206 2018 2019 1561
 2017 2016 1207 2020 1205 1278 1279
 2084 2085 222 1201 1281 1014 203 1209
 1208 489 223 221 210 209 208 1562 2355
 497 213 216 211 1563 1573 1552 1578
 2369 492 1554 2371 2100 2102 490 220
 503 1580 507 807 2348 808 799 374 1868
 373 797 378 802 377 376 385 795 231 375
 379 380 1332 1333
 2147 1575 509 1313 2130 1311 212 493
 215 199
 2101 798 372 801 806 1320 2129 495
 2118 784
 1395 804 1331 1318 2112 2113 197 800
 1389 1396
 788 227 1310 226 2115 1308 2111 198
 384 805
 1386 787 369 1314 1406 1407 2217 793
 1402 1403

LIGAND CHEMISTRY ANALYSIS
(SKF 108738):-

Donor atoms: 1 6 11 21 33 40

No donor atoms 6

Acceptor atoms: 4 9 21 25 36 43

DOCKING HIV-1 PROTEASE WITH SKF 108738Doing GA no population(s) 3 size 100
selection pressure 1.100000

Operation (vdw_ext)	Fitness S (hb_int)	S (hb_ext)	S (vdw_int)	S	S
0	43.91	15.04	53.95		
0.00	-226.51				
497	45.27	12.35	50.35	0.00	-
177.71					
532	46.09	4.95	49.72	0.00	-
133.25					

Conclusion:

BEST DOCKING COMPOUND

“SKF 108738”**Acknowledgement:**

We would like to express deep sence of gratitude to Mr. Sanjaykumar Choubey for his guidance, persuasion, sincere work, and we also thankful to Mr. Pramod Deshmukh for his valuable suggestion and enthusiasm, which helped us a lot.

Reference:

- Davies DR (1990). "The structure and function of the aspartic proteinases". *Annu Rev Biophys Chem* **19**: 189–215. doi:10.1146/annurev.bb.19.060190.001201. PMID 2194475.
- Brik A, Wong CH (January 2003). "HIV-1 protease: mechanism and drug discovery". *Org. Biomol. Chem.* **1** (1): 5–14. doi:10.1039/b208248a. PMID 12929379.
- Kräusslich HG, Ingraham RH, Skoog MT, Wimmer E, Pallai PV, Carter CA (February 1989). "Activity of purified biosynthetic

proteinase of human immunodeficiency virus on natural substrates and synthetic peptides". *Proc. Natl. Acad. Sci. U.S.A.* **86** (3): 807–11. doi:10.1073/pnas.86.3.807. PMID 2644644.

- Kohl NE, Emini EA, Schleif WA, *et al.* (July 1988). "Active human immunodeficiency virus protease is required for viral infectivity"
- Seelmeier S, Schmidt H, Turk V, von der Helm K (September 1988). "Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A". *Proc. Natl. Acad. Sci. U.S.A.* **85** (18): 6612–6. doi:10.1073/pnas.85.18.6612. PMID 3045820.
- McPhee F, Good AC, Kuntz ID, Craik CS (October 1996). "Engineering human immunodeficiency virus 1 protease heterodimers as macromolecular inhibitors of viral maturation". *Proc. Natl. Acad. Sci. U.S.A.* **93** (21): 11477–81.
- Jaskólski M, Tomasselli AG, Sawyer TK, Staples DG, Heinrikson RL, Schneider J, Kent SB, Wlodawer A (February 1991). "Structure at 2.5-Å resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor". *Biochemistry* **30** (6): 1600–9. doi:10.1021/bi00220a023. PMID 1993177.
- Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SB, Wlodawer A (December 1989). "Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution". *Science (journal)* **246** (4934): 1149–52. PMID 2686029. <http://www.sciencemag.org/cgi/pmidlookup?vie w=long&pmid=2686029>.
- F.K. Brown (1998). "Chapter 35. Chemoinformatics: What is it and How does it Impact Drug Discovery". *Annual Reports in Med. Chem.* **33**: 375. doi:10.1016/S0065-7743(08)61100-8.
- Brown, Frank (2005). "Editorial Opinion: Chemoinformatics – a ten year update". *Current Opinion in Drug Discovery & Development* **8** (3): 296–302.
- Gasteiger J.(Editor), Engel T.(Editor): *Chemoinformatics : A Textbook*. John Wiley & Sons, 2004, ISBN 3-527-30681-1
- A.R. Leach, V.J. Gillet: *An Introduction to Chemoinformatics*. Springer, 2003, ISBN 1-4020-1347-7