AcademicPres

Available online at www.notulaebotanicae.ro

Print ISSN 0255-965X; Electronic ISSN 1842-4309

Not. Bot. Hort. Agrobot. Cluj 37 (2) 2009, 264-269

Notulae Botanicae Horti Agrobotanici Cluj-Napoca

Computational Analysis of POL Gene of Human Immunodeficiency Virus Type 1 (HIV-1)

Pabitra M. BEHERA[1)], Bijay K. SETHI[2)], Kambaska K. BEHERA[2)] Debashrita PANI[2)], Santilata SAHOO[2)] Sudarshan PADHY[1)]

[1)] *Utkal University, Department of Mathematics Statistics and Computer Science, Vani Vihar, Bhubaneswar-751004, India; kambaska@yahoo.co.in*

[2)] *Utkal University, Department of Applied Microbiology, Vani Vihar, Bhubaneswar-751004, India; kambaska@yahoo.co.in*

Abstract

AIDS results from an infection with HIV-1 virus. It is a class of retrovirus whose genome contains the genes for reverse transcriptase. The gene that codes for reverse transcriptase in HIV-1 is the POL gene. POL genes are found in many retroviruses, including a number that are harbored by human populations. In this paper POL genes of HIV-1 virus were studied using bioinformatics tools. It was observed that there exists a close resemblance among various strains endemic in Africa, China and India. In addition, they express a wide range of functional proteins after infection.

Keywords: AIDS, HIV-1, POL gene, retrovirus

Introduction

HIV is thought to have originated in non-human primates of Africa and transferred to humans early in the 20th century (Worobey *et al.*, 2008). Generally two species HIV-1 and HIV-2 are responsible for human infection. HIV-1 is thought to have originated in southern Cameroon after jumping from wild chimpanzees (*Pantroglodytes*) to humans during the twentieth century (Gao *et al.*, 1999) (Keele *et al.*, 2006). It evolved from a Simian Immunodeficiency Virus (SIVcpz) [4]. HIV-2, on the other hand, may have originated from the Sooty Mangabey (Cercocebusatys), an Old World monkey of Guinea-Bissau, Gabon, and Cameroon (Reeves *et al.*, 2002). HIV was discovered by the French scientist Luc Montagnier in 1983 and American researcher Robert Gallo in 1984, but Luc Montagnier is given the credit for discovering it. The unique structure of HIV makes it different from other leading retroviruses. It is around 120 nm in diameter and roughly spherical. HIV-1 is composed of two copies of single-stranded RNA enclosed by a conical capsid comprising the viral protein p24 (Fig. 1.). The RNA is 9749 nucleotides long (Ratner *et al.*, 1985) and surrounded by a plasma membrane. The single-strand RNA is tightly bound to the nucleocapsid proteins, p7, p6 and enzymes like reverse transcriptase and integrase. This nucleocapsid protects the RNA from digestion by nucleases. A matrix composed of an association of the viral protein p17 surrounds the capsid, ensuring the integrity of the virion particle. The nucleocapsid is then covered by a thick spike like envelop comprising glycoproteins gp120 and gp41. The Cryo-electron microscopy reveals that three copies of gp120-gp41 heterodimers form a trimer as the envelope spike (Zhu *et al.*, 2006). This is only hypothetical as there are various possibilities mentioned in publications from different sources. Similar to leading retroviruses HIV has several structural genes coding for structural proteins. Besides these structural genes it also has nonstructural genes (accessory genes) for unique functionality. The GAG gene provides the basic physical infrastructure of the virus, and POL provides the basic mechanism by which retroviruses reproduce, while the others help HIV to enter the host cell and enhance its reproduction .

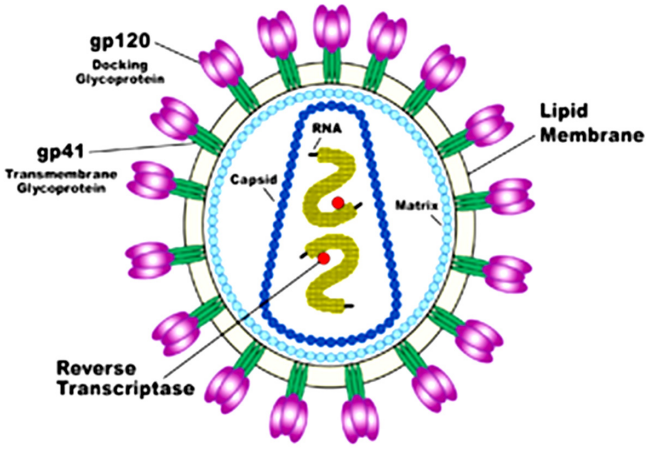

Fig. 1. Diagram of HIV retrovirus showing various GAG and POL proteins responsible for viral structure and function.

Earlier, based on comparisons of HIV sequences from various sources, the origin of the AIDS virus can be pin-

pointed to Africa in the 1930's. These comparisons reveal that the earliest sample of HIV is from a blood sample drawn from an African in 1959. This paper reveals the current status of HIV-1 infection reported from Africa, China and India and their evolutionary relationship based on the sequence comparison, structural biology of POL gene retrieved from various databases.

Materials and methods

Materials

Nucleotide and Protein sequence database of NCBI, EMBL and DDBJ now contain vast data pertaining to HIV-1 infection. Their size increases significantly due to rapid submission of raw sequences by researchers from various laboratories over the globe. This trend will continue until there is funding for sequencing projects in molecular biology. As an initial approach to ascertain the current status of HIV-1 infection in Africa, China and India the Nucleotide and protein databases of NCBI was searched with the queries "HIV1 pol Africa", "HIV1 pol China", "HIV1 pol India" respectively. It generated 10084 nucleotide and 9513 protein entries. (Tab. 1) Out of total entries retrieved for each query, the first 10 entries were taken for comparative analysis. Out of 10 nucleotide sequences obtained from Africa, the last four were dropped from analysis as they were reported to be complete genome entries. Similarly for proteomics of HIV-1 the protein structures reported to date was searched in RCSB, PDB (Protein Data Bank) with the query "HIV1 pol". The search generated four structure hits (1Z1H, 1Z1R, 2BB9, 2NPH) and four ligand hits (AKC, HBB, HBH, SO4). The PDB files for four reported structures and their FASTA (Pearson, 1990) formatted sequences were downloaded for analysis.

Tab. 1. Current status of HIV-1 POL gene reported from Africa, China and India (Data retrieved from NCBI)

Type of sequences	Gene name	Country	Entries
Nucleotide sequences			
	HIV-1 POL	Africa	3157
	HIV-1 POL	China	5513
	HIV-1 POL	India	1414
	Total		10084
Protein sequences			
	HIV-1 POL	Africa	3047
	HIV-1 POL	China	5175
	HIV-1 POL	India	1291
	Total		9513

Methods

The FASTA formatted files for nucleotide and protein sequences were prepared for upload to nucleotide tools and protein tools of SDSC (San Diego Supercomputer Centre) Biology Workbench 18 for genomic and proteomic analysis. A session was opened with login id and password to and sequences were uploaded for analysis in Workbench chest.

Viral Genomics

The 26 nucleotide sequences were selected from the session body and threaded to CLUSTALW tool (Thompson *et al.*,1994) for Multiple Sequence Alignment (MSA) (Lipman *et al.*,1989) with parameters reading (DNA transitions weight: 0.5, Weight matrix: ClustalW (1.6), Use negative matrix: No, Gap open penalty: 15.00, Gap extension penalty: 6.66, Delay divergent sequences: 30). The result alignment was then analyzed in DRAWGRAM tool (Felsenstein, 1989) (parameters reading Short Labels: No, Tree grows: Horizontally, Tree style: Cladogram, Use branch lengths: Yes, Scale of branch length: Automatically rescaled, Specified cm/unit branch length: 5.0, Angle of labels: 90, Depth/Breadth of tree: 0.7, Stem length/tree depth: 0.05, Character height/tip space: 0.3333, Ancestral nodes: default, Vertical margin (cm): 1.0, Horizontal margin (cm)>: 1.0) to obtain rooted phylogenetic tree.

Viral Proteomics

The 30 protein sequences were selected from the session body and threaded to CLUSTALW tool for Multiple Sequence Alignment (MSA) with parameters reading. (Weight matrix: PAM series, Use negative matrix: No, Gap open penalty: 10.00, Gap extension penalty: 0.02, Delay divergent sequences: 30, Residue-specific gap penalties: On, Hydrophilic gap penalties: On, Hydrophilic residues: GPSNDQEKR, Gap separation distance: 0, End gap separation penalty: Off) The result alignment was then analyzed in DRAWGRAM tool (parameters reading Short Labels: No, Tree grows: Horizontally, Tree style: Cladogram, Use branch lengths: Yes, Scale of branch length: Automatically rescaled, Specified cm/unit branch length: 5.0, Angle of labels: 90, Depth/Breadth of tree: 0.7, Stem length/tree depth: 0.05, Character height/tip space: 0.3333, Ancestral nodes: default, Vertical margin (cm): 1.0, Horizontal margin (cm)>: 1.0) to obtain rooted phylogenetic tree. The PDB files (1Z1H, 1Z1R, 2BB9, 2NPH) were studied in Discovery Studio 22 in order to obtain a clear confirmation of chains, water molecules and hetero atoms and their secondary structures. Then docking of ligand entries (AKC, HBB, HBH) on proteins (2BB9, 1Z1H, 1Z1R) were studied in GOLD 3.0.1 (Jones *et al.*, 1997) (Jones *et al.*, 1995) (Nissink *et al.*, 2002) (Verdonk *et al.*, 2003) with appropriate parameters (Waters: 0, Metals: 0, Set atom types: Ligand, Early termination is on if top 3 solutions are within 1.5 Å R.M.S.D, Define site using: Ligand, Active site radious: 10.0, Detect Cavity: on with default GOLD Score and Genetic Algorithm parameters). Finally the GOLD run results were studied in Silver 1.1 (Verdonk *et al.*, 2003) to find the interactions of li-

gands (three best conformations) with amino acid residues of proteins respectively.

Results and discussion

Viral Genomics

The phylogenetic analysis based on Multiple Sequence Alignment of nucleotide sequences reported from Africa, China and India (Fig. 2.) reveals that there exists a close resemblance among the viral strains. The India strains are more resembling to African strains than the strains from China. Out of 26 strains studied a unique strain from China (AF395547) is closely related to African strain (AY461492).

Viral Proteomics

The phylogenetic analysis based on Multiple Sequence Alignment of protein sequences reported from Africa, China and India (Fig. 3.) reveals that there exists an anomalous resemblance among the viral strains. This may be due to various expression features of the proteins under different physiological conditions.

The viral proteins reported from PDB (1Z1H, 1Z1R, 2BB9, 2NPH) were found to be of "Hydrolase" type.

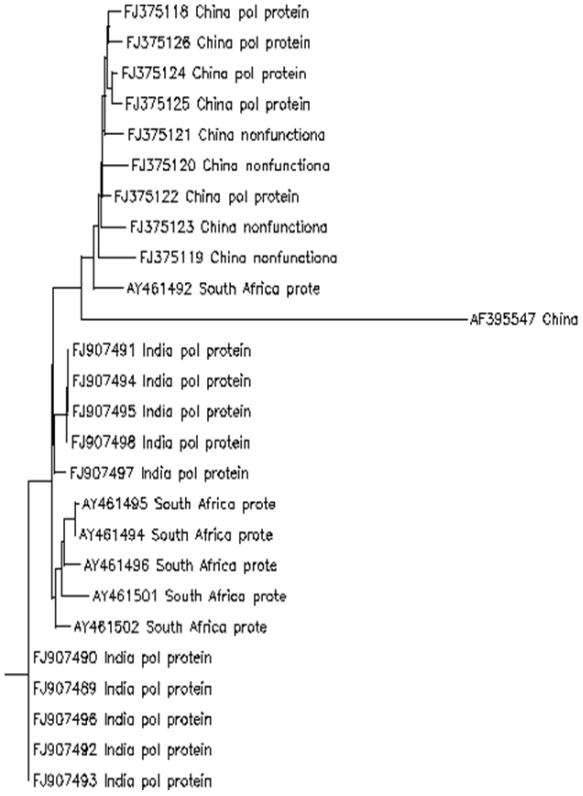

Fig. 2. Rooted phylogenetic tree obtained from DRAWGRAM tool of SDSC (San Diego Supercomputer Centre) Biology Work Bench on the Multiple Sequence Alignment of 26 neucleotide sequences from Africa, China and India respectively

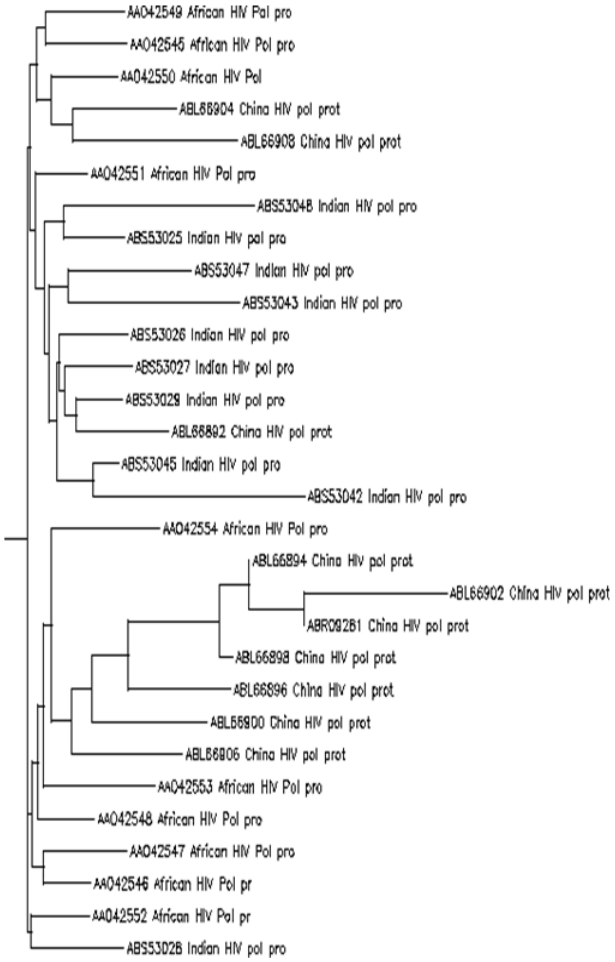

Fig. 3. Rooted phylogenetic tree obtained from DRAWGRAM tool of SDSC (San Diego Supercomputer Centre) Biology Work Bench on the Multiple Sequence Alignment of 30 protein sequences from Africa, China and India respectively

The structural analysis illustrated that the amino acid chain of 99 residues (Chain A) and 98 residues (Chain B) were conserved among these four models. Two amino acid chains of 4 residues (Chain S) and 6 residues (Chain T) were unique features of 2NPH. The GOLD docking results were studied in SILVER with most favorable conformations of three ligand entries on three protein models as discussed above. Then a protein subset was defined by assigning residues that are within 4Å areas of ligands. Residues that are out of the subset were dropped from the view and final interactions were studied by generating hydrogen bonds. On this aspect the docking of AKC on 2BB9 shows hydrogen bonding with residues (ASP25, ASP29, ASP225, ASP230, GLY27 and ILE50) (Fig. 4.). Similarly docking of HBB on 1Z1H shows hydrogen bonding with residues (ASP29, ASP125 and GLY27) (Fig. 5.) and docking of HBH on 1Z1R shows hydrogen bonding with residues (ASP29, ASP30, ASP125, and GLY27) (Fig. 6.).

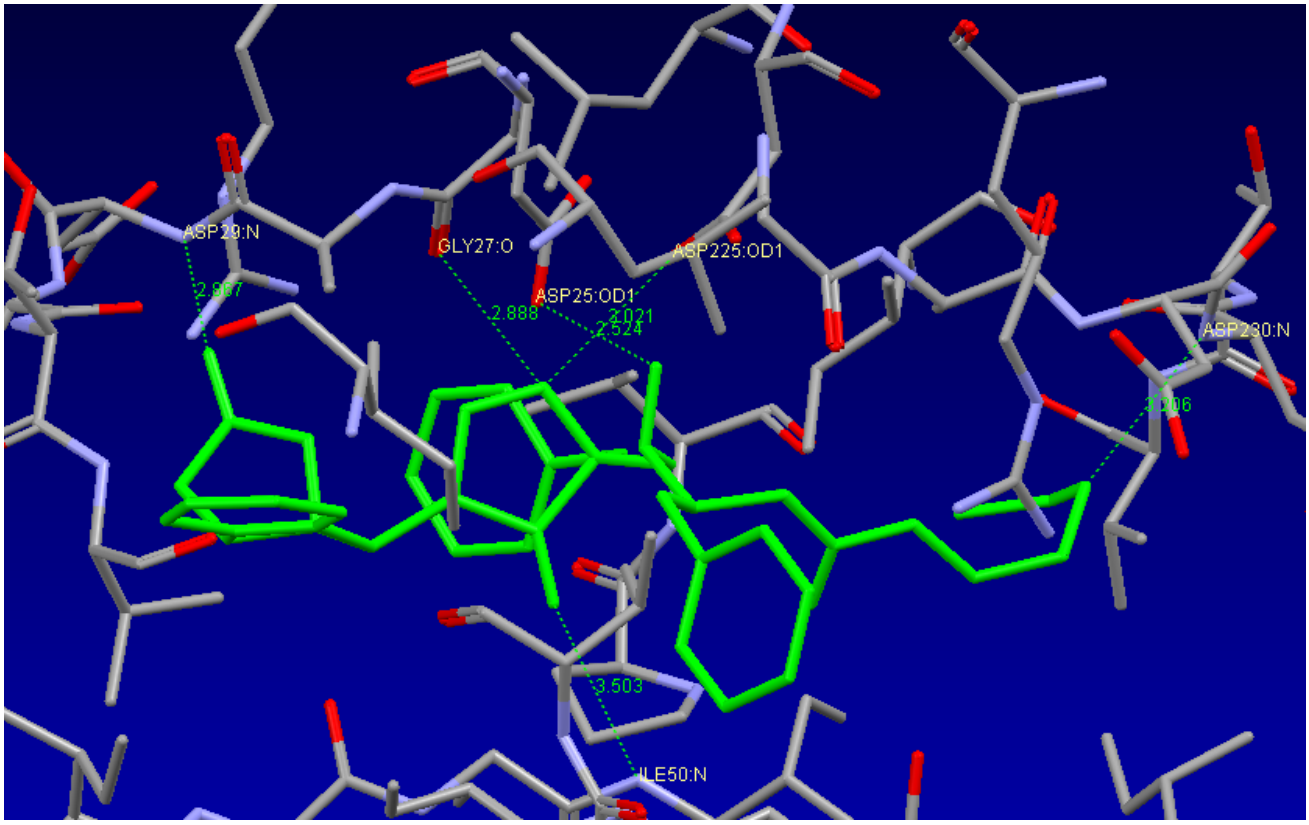

Fig. 4. The docking of AKC on 2BB9 showing hydrogen bonding with residues ASP25, ASP29, ASP225, ASP230, GLY27and ILE50.

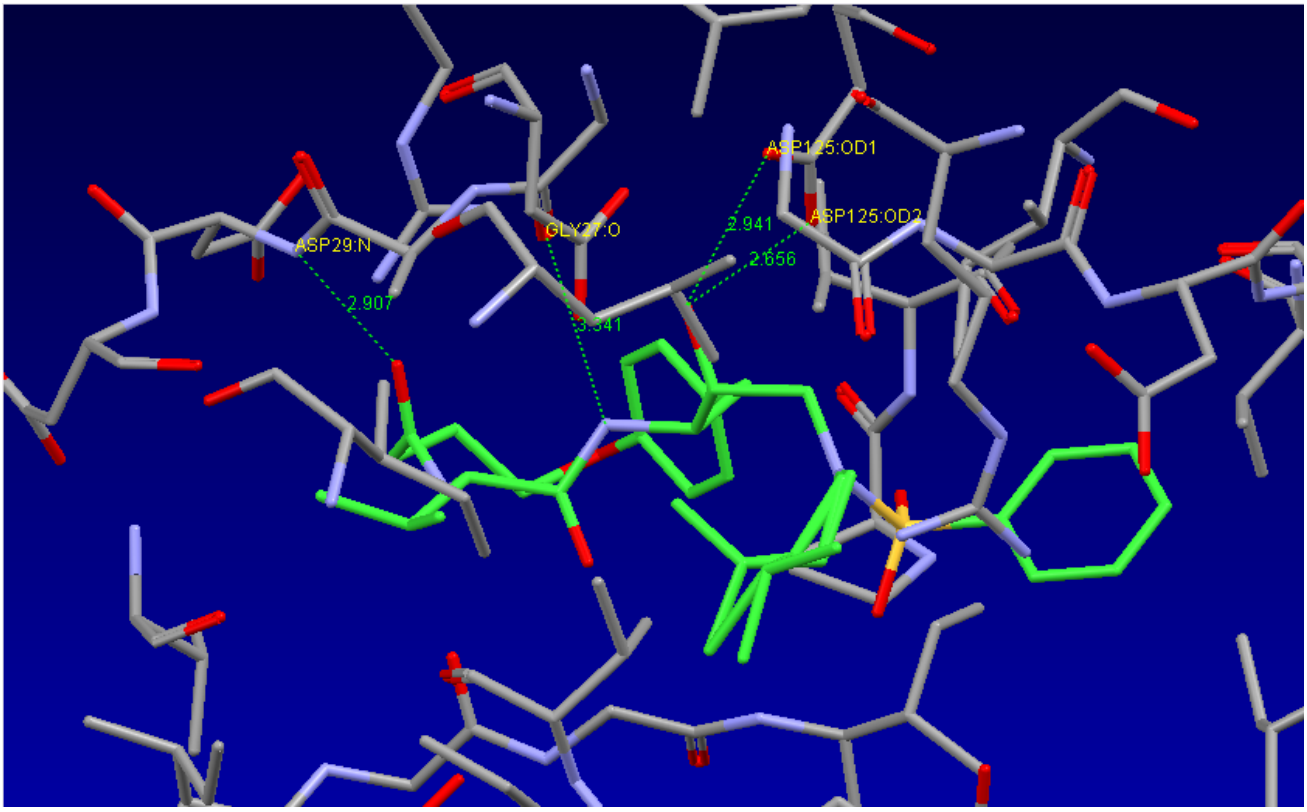

Fig. 5. The docking of HBB on 1Z1H showing hydrogen bonding with residues ASP29, ASP125 and GLY27.

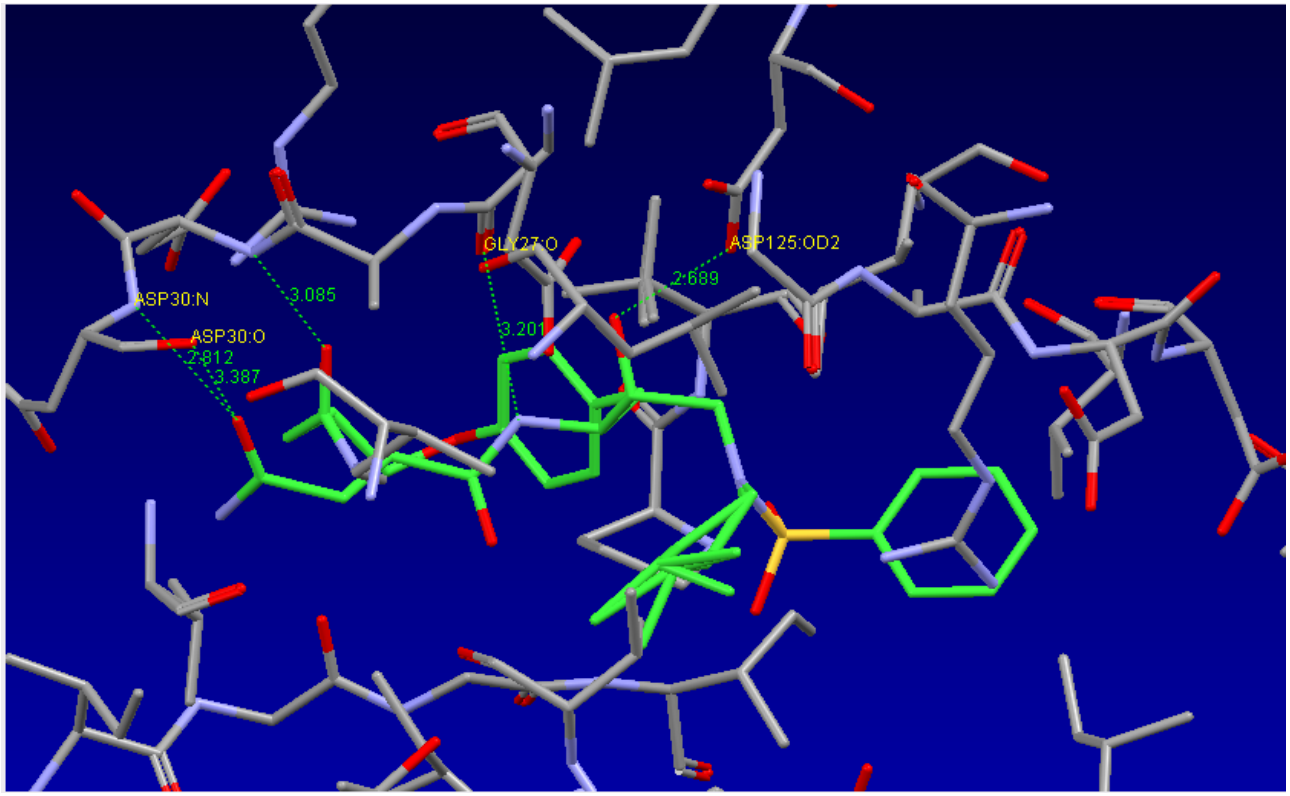

Fig. 6. The docking of HBH on 1Z1R showing hydrogen bonding with residues ASP29, ASP30, ASP125 and GLY27.

Conclusions

Computational analysis of POL gene of Human Immunodeficiency Virus Type 1(HIV-1) shows that there exists a close resemblance among various strains that are endemic in Africa, China and India. Again, they express a wide range of functional proteins after infection. A further study includes the QSAR of all available potential ligands and their interaction on various expression patterns of the pol gene.

References

Worobey, M., M. Gemmel and D. E. Teuwen (2008). Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature. 455 (7213):661-664.

Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp and B. H. Hahn (1999). Origin of HIV-1 in the Chimpanzee Pan troglodytes troglodytes. Nature. 397 (6718):436–441.

Keele, B. F., F. Van Heuverswyn, Y. Y. Li, E. Bailes, J. Takehisa, M. L. Santiago, F. Bibollet-Ruche, Y. Chen, L. V. Wain, F. Liegois, S. Loul, E. Mpoudi Ngole, Y. Bienvenue, E. Delaporte, J. F. Y. Brookfield, P. M. Sharp, G. M. Shaw, M. Peeters and B. H. Hahn (2006). Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1. Science Online. 2006-05-25:523.

http://www.nature.com/nature/journal/v397/n6718/abs/397436a0_fs.html

Reeves, J. D. and R. W. Doms (2002). Human Immunodeficiency Virus Type 2. J. Gen. Virol. 83 (Pt 6):1253-1265.

Brown, Phyllida (25 May 1991). The strains of the HIV war. New Scientist. http://www.newscientist.com/channel/health/mg13017703.800-the-strains-of-the-hiv-war.html.

Nobel Prize awarded for AIDS, cervical cancer research. Los Angeles Times. http://articles.latimes.com/2008/oct/06/science/sci-nobel7. Retrieved on 2008-10-06.

Ratner, L., W. Haseltine and R. Patarca (1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature. 313(6000):277-284.

Zhu, P., J. Liu and J. Jr. Bess (2006). Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature. 15:817-818.

Introduction to HIV Sequence Compendium in (2008). annual review. http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2008/frontmatter.pdf

Centers for Disease Control and Prevention. Pneumocycstis Pneumonia-Los Angeles. Morbidity and Mortality Weekly Report. 30:250-252. http://www.cdc.gov/hiv/resources/reports/mmwr/pdf/mmwr05jun81.pdf.

Centers for Disease Control and Prevention. Kaposi's Sarcoma and Pneumocycstis Pneumonia among Homosexual Men - New York City and California. Morbidity and MortalityWeeklyReport.30:305-308. http://www.cdc.gov/

hiv/resources/reports/mmwr/pdf/mmwr04jul81.pdf.

The National Centre for Biotechnology Information, http://www.ncbi.nlm.nih.gov/.

The European Molecular Biology Laboratory, http://www.embl-heidelberg.de/.

The DNA Data Bank of Japan, http://www.ddbj.nig.ac.jp/.

The Research collaborator for Structural Bioinformatics, Protein Data Bank. http://www.rcsb.org/pdb/home.

Pearson, W. R. (1990). Meth. Enzymol. 183:163-198. 1990.

The Biology Workbench of SDSC (San Diego Supercomputer Centre), http://workbench.sdsc.edu/

Thompson, J. D., D. G. Higgins, T. J. Gibson and W. Clustal (1994). Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res. 22:4673-4680.

Lipman, D. J., S. F. Altschul and J. D. Kececioglu (1989). Proc, Natl. Acad. Set. USA. 86:4412-415.

Felsenstein, J. (1989). PHYLIP-Phylogeny Inference Package (Version 3.2). Cladistics 5:164-166.

Accelrys, D. S. V1.7. Discovery Studio is flexible software environment for bioinformatics and drug discovery. (Accelrys Software Inc).

Jones, G., P. Willett, R. C. Glen, A. R. Leach and R. Taylor (1997). Development and Validation of a Genetic Algorithm for Flexible Docking. J. Mol. Biol. 267:727-748.

Jones, G., P. Willett and R. C. Glen (1995). Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. J. Mol. Biol. 245:43-53.

Nissink, J. W. M., C. Murray, M. Hartshorn, M. L.Verdonk, J. C. Cole and R. Taylor (2002). A New Test Set for Validating Predictions of Protein-Ligand Interactions. Proteins. 49(4):457-471.

Verdonk, M. L., J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor (2003). Improved Protein-Ligand Docking using GOLD. Proteins.