Vol. 1, No. 1 January - June 2010 ISSN 2072-4659

NUST JOURNAL OF NATURAL SCIENCES

National University of Sciences & Technology
Islamabad - Pakistan

NUST JOURNAL OF ENGINEERING SCIENCES - Dec. 2009, Vol. 2, No. 1

NUST
Publishing
National University of Sciences & Technology
Sector: H-12, Islamabad-44000, Pakistan
E-mail: dgpub@nust.edu.pk
Website: www.nust.edu.pk

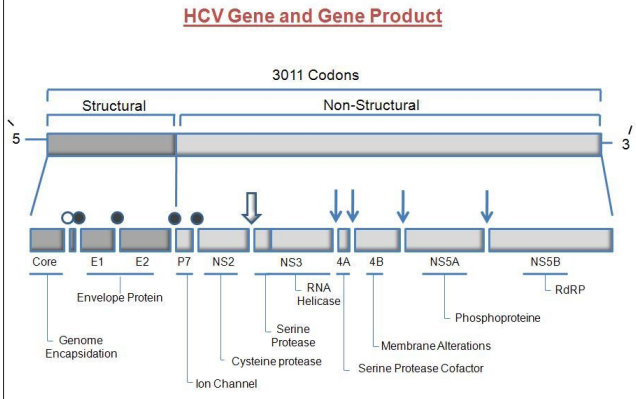

Fig. 1. Proteins made by HCV RNA and their cleavage patterns. Closed circles refer to signal peptidase cleavage sites; the open circle refers to the signal peptide peptidase cleavage site.

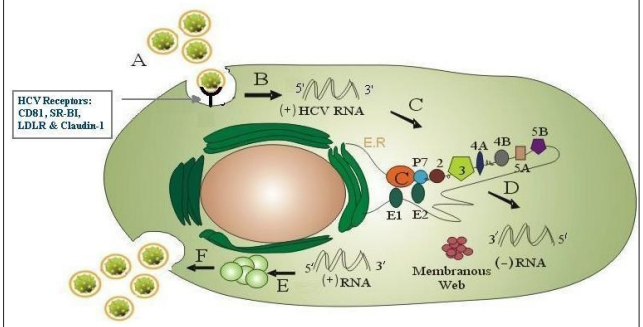

Fig. 2. (A) Recepter (CD81, SR-BI, LDLR and claudin-1) mediated entry, cytoplasmic release and uncoating the HCV (B) IRES-mediated translation and processing of the polyprotein (C) HCV RNA replication (D); packaging and assembly (E); virion maturation (F) Release of the virus

Bacteriophages; Major Applications and Phage Therapy

Waqas Nasir Chaudhry, Irshad ul Haq, Saba Safdar, Ishtiaq Qadri
NUST Center of Virology & Immunology (NCVI),
National University of Sciences & Technology (NUST) Islamabad, Pakistan.
waqas.mmg@gmail.com

Abstract

An introduction to bacteriophages along with the general areas of application is described with special and detailed discussion of therapeutic use of phages and phage enzymes. Phage therapy being safer and faster than conventional chemotherapy has been employed for a while but its incorporation into general disease combating methods has to make a few pit stops those being the proper understanding of phage behaviour in vivo among others. The experiments done on model organisms and humans are described. Industrial benefits of phages, their surface cleaning abilities as well as use of bacteriophages in livestock and poultry development are highlighted. Phages as therapeutics hold a promising future due to their novel nature and mechanism of action for an underdeveloped country like ours.

Key words: Bacteriophage, Phage therapy, Antibiotics, lysin, Inhibition.

Introduction

Bacteriophage or phage (short for from Greek "phagein" that means "to eat" or "to nibble") are highly specific viruses that invade bacterial cells. They are obligate intracellular parasites that multiply inside bacteria by making use of some or all of the host biosynthetic machinery (i.e., viruses that infect bacteria) ultimately resulting in death of bacteria.

First reported discovery of viruses infecting bacteria was by Twort in the year 1915. He studied excretory samples of some patients of diarrhea and was able to isolate a peculiar filterable infectious agent from the samples. This discovery was further validated when in 1917 d'Hérelle also reported a similar finding. This discovery occurred about 20 years before practical application of penicillin, the first antibiotic. At that time phage therapy was considered as possible treatment method against bacterial infections (Ho K 2001). Although phage therapy is being practiced in Soviet unions and Eastern Europe (Chanishvili *et al.* 2002) but it was abandoned by West in 1940s with the discovery of antibiotics. However antibiotic resistance emergences in bacteria i.e. multi drug resistant (MDR) and extremely drug resistant bacteria (XRD) are major developing threats that have recently motivated the western scientific community to reevaluate phage therapy for bacterial infections that are incurable by conventional chemotherapy (Shigenobu *et al.* 2005).

Phage as tools in molecular biology

Phage research has had a key impact on molecular biology. The T series of bacteriophages had a central role in the development of molecular biology. In their book *Phage and the Origins of Molecular Biology*, Cairns *et al.* show how phages have contributed not only to the understanding of essential cellular processes, but also to the development of a considerable number of important genetic and biochemical tools. For example, the realization that viable bacteriophage lambda particles could be constructed with a significant portion of their genome deleted led to the development of insertional and replacement vectors, as well as cosmids and integrative plasmids. Phage serine integrates, particularly those of *Streptomyces* phage C31 have been exploited by Michele P Calos (Stanford University, USA) to integrate foreign DNA into mammalian cells and Drosophila (Groth *et al.* 2004) with the goal of producing transgenic animals or curing biochemical defects (Ginsburg *et al.* 2005), (Held *et al.* 2005). In addition, phage packaging signals, promoters and terminators, together with a great variety of enzymes, are used in today's molecular biology laboratory including polynucleotide kinases, DNA ligases, DNA polymerases, RNA polymerases, recombinases, single stranded DNA binding proteins, endo and exonucleases, and even methylases and restriction endonucleases (Roberts RJ *et al.* 2003).

Medical applications and industrial applications

Phage therapy is an area of phage applications that is being explored extensively due to emergence of antibiotic resistant bacterial strains. Specific phages can be administered to infected individuals that can lyse the bacteria thus ridding the patient of infection. Proteins and antibodies are purified by phage display technology (Zilka *et al.* 2003). These can then be used as therapeutics that would act either as agonists or through the inhibition of receptor ligand interactions. Another important use of phage is in diagnosing bacterial infections and in epidemiological studies. Phage typing is a technique used to diagnose bacterial infection by using phages. Phage typing is also used in epidemiological studies i.e. to identify if two epidemics are caused by the same bacterial strain or not (Sharp *et al.* 2001). Phages are used in various diagnostic procedures for example, determining antibiotic sensitivity for slow growing Mycobacterium tuberculosis (M Pai *et al.* 2005). Phages can also be tagged with a flourescein dye like antibodies and used in diagnostic; the tag can be detected using microscopy techniques (Hennes

Corresponding Author: Waqas Nasir Chaudhry

et al. 1995).

Bacteriophages are used in food processing industries to avoid bacterial contamination of food as well as eradication of biofilms which in turn increases shelf life of products. Bacterial contamination in food industries is a hazard that causes the industry immense damage. The most damaging bacterium of food industry is *Listeria monocytogenes* and phages against this are used in production and processing of food. Sometimes foods need a threshold level of bacteria for the enzymes to work, anything more than that value can be poisonous for the enzymes used in the process. Phages are used to control these bacterial loads (Flaherty *et al.* 2001), (Leverentz *et al. 2001*). Bacteria exist as biofilms in nature and these biofilms act as a shield to antimicrobials and biocides. Surfaces that are in contact with food can be decontaminated of any bacterial colonization using phages (Bassett *et al.* 2007). Phages are used to take out these biofilms so that the antimicrobials and biocides can effectively remove the contaminants. Bacteriophages are also used in mining industries to assist bacteria in different steps of metal and coal processing. A bacteriophage called phi Ac1 has been reported to assist bacteria in mining operations (Ward *et al.* 1992) which help bacteria remove sulphur from coal prior to burning.

Phages in agriculture & live stock

Phages are used in agricultural settings to control plant pathogenic bacteria by introducing specific phages to the fields (J.B. Jones *et al.* 2007). Fire blight in apple trees as well as Tomato and Pepper spots are targets of phage therapy (Gill JJ *et al.* 2003). Also phage systems like Cre lox P can be used to generate transgenic plants. Phage can also play a very important role in the decontamination of meat by treating the animal with phages before its slaughter (Sklar *et al.* 2001)

Antibiotic resistance and future problems

S. aureus is not the only problem, the CDC estimates that in some areas, 30% of pneumonia caused by *Streptococcus pneumoniae* is resistant to penicillin, whereas virtually all cases were susceptible in the 1970s. Vancomycin started failing to keep some *Enterococcus* (*faecium and faecalis*) infections in check in late 1988, necessitating new aggressive combination regimens. By 1993, according to the NIH, more than 10% of hospital acquired enterococci infections reported to the CDC were attributed to vancomycin resistant *Enterococcus* (VRE) *faecium*.

Aventis's (Strasbourg, France) drug Synercid was introduced in 1999 as a new weapon against VRE, but some resistance was observed before it even reached the market (G.M. Eliopoulos *et al.* 1998). Figure 1.

Phage therapy

Phage therapy is in fact the use of specific bacteriophages to kill pathogenic bacterial strains. Classically it implies the use of whole phage that infects bacterial cells and causes its lysis, however recently phage Lytic enzymes are being used to cause lysis of bacterial cells. Whole phage

Antimicrobial	MRSA n = 82	MSSA n = 18
Vancomycin	0%	0%
Clindamycin	73.2%	66.7%
Erythromycin	90.2%	66.7%
TMP/SMX[3]	42.7%	22.2%
Ciprofloxacin	79.3%	72.2%
Tetracycline	87.8%	77.7%

Fig. 1. MRSA=Methicillin-resistant S. aureus, MSSA= Methicillin-susceptible S. aureus, P/SMX=Trimethoprim/ sulfamethoxazol. Antimicrobial resistances of MRSA and MSSA isolated from infections in ICU (Aysen et al. 2006).

infects bacterial cell, multiplies and causes lysis of the cell thus releasing large number of phage progeny. These newly released phages further infect bacterial cells and radically decrease the bacterial load in the infected organism. Phages can be administered orally, topically or directly into tissues via injections. Besides using phages to cause bacterial lysis, phages can be used to deliver non phage genes to infected cells coding for antimicrobials (West water *et al.* 2003). Figure 2.

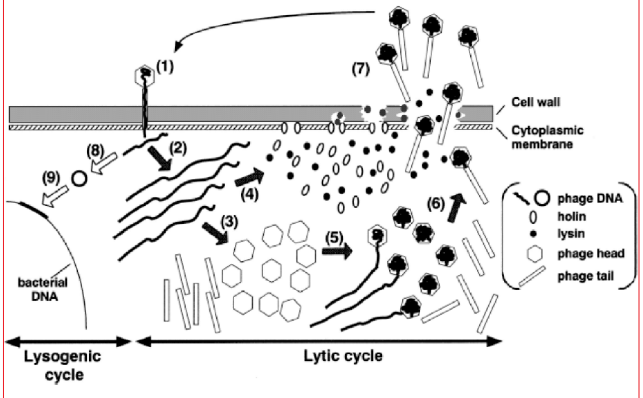

Fig.2. Schematic illustration of phage induced bacteriolysis. (1) Adsorption and DNA injection; (2) DNA replication; (3) production of head and tail; (4) synthesis of holin and lysin; (5) DNA packaging; (6) completion of phage particle; (7) disruption of the cell wall and release of the progeny; (8) circularization of phage DNA; (9) integration of the phage DNA into the host genome. (Shigenobu et al.2005).

Work on efficacy of phage therapy in living organisms was initiated in 1980's by Smith *et al.*via experimentation in veterinary animals. According to Smith's experiments, one dose of phage either intramuscular or intracerebral is far more effective against *E.coli* K1 infection in mice as compared to treatment with antibiotics like ampicillin, chloramphenicol or trimethoprim. This opened the way for further experiments on model organisms against a wide range of bacterial infections like *E. coli*, (Merril CR *et al.* 1996), (Chibani Chennoufi S *et al.* 2004) *Pseudomonas aeruginosa* (Ahmad SI *et al.* 2002), *Acinetobacter baumanii* (Soothill JS. *et al.* 1992), *Klebsiella pneumoniae*, *Enterococcus faecium* (vancomycin resistant strain, VRE),

(Biswas B *et al.* 2000) *Vibrio vulnificus*, (Cerveny KE *et al.* 2002) and *Salmonella* sp.

Staphylococcus aureus is a pathogen responsible for inflammatory diseases, toxic shock syndrome and food poisoning. Some of the antibiotic resistant strains of *S. aureus* include the methicillin resistant *staphylococcus aureus* (MRSA) (Hiramatsu K *et al.*2001), (Shimada K *et al.* 2004) and vancomycin resistant *S. aureus* (VRSA) (Chang S *et al.*2003), (Kacica M. *et al.* 2004). The drug resistant strains of this specie are rapidly on rise and efforts must be devoted to combat this bacterial dissemination through phage therapy. Also *S. aureus* strains that are resistant to a relatively new antibiotic linezolid are already being reported in USA and Europe (Pillai SK *et al.* 2002). According to the CDC, methicillin resistant *Staphylococcus aureus* (MRSA) accounted for nearly 60% of nosocomial *S. aureus* infections in 2001, a figure that had nearly doubled over the previous decade. Although most MRSA can still be treated with powerful antibiotics, some super bugs can shrug off even the strongest drugs. The first reported case of resistance to Pfizer's (Groton, CT, USA) Zyvox, the last line of defense against MRSA, was reported a little more than a year after the drug was approved. (S. Tslodras *et al.* 2001).

Pseudomonas aeroginosa is rapidly becoming resistant to antibiotics. Regarding this bacterium, Yun *et al's* recent work is worth mentioning. In their work, they identified two phages (MPK1 and MPK6) against *pseudomonas aeroginosa* strain PAO1. In experiments on mouse models, MPK1 administration; and to a lesser extent MPK6 administration decreased the mortality rate associated with PAO1 induced peritonitis sepsis. Also the mice treated with either one of these bacteriophages had lesser bacterial loads in livers, lungs and spleens. Another animal model used was Drosophila melanogester. Both of the phages were introduced via feeding and resulted in delayed PAO1 induced killing of D. melanogester. This experiment further substantiated the efficacy of MPK1 and MPK6 for *pseudomonas aeroginosa* infections.

Another important practical example of phage therapy usage was presented by Naka *et al.*who saved a large number of fish infected by *Lactococcus garvieae* and *Pseudomonas plecoglossicida* (Nakai *et al.* 1999), (Park *et al.* 2003). Phages have also been reported successful in fight against food poisoning in eliminating pathogens like *salmonella sp*. (Goode *et al.* 2003), *Campylobacter jejuni*, and *Listeria monocytogene*, ((Leverentz *et al.* 2004).

Research is also being done on the statistical and mathematical aspects of phage therapy to better elucidate phage host interaction mechanisms and if a threshold quantity of host cells are required for viable infection with phages or not (Payne *et al* .2000-2002). A hand wash solution enriched with phages has been reported to reduce staphylococcal contamination by 100 fold as compared to a hand wash solution that is phage free. In another experiment, application of a staphylococcal phage prevented abscess formation in rabbit model of wound infection (Wills *et al* .2005).

Using phage lytic enzymes

The mechanism of phage infection involves the progeny release step which is necessary for viability of infection. In order to release the newly formed phage particles, bacteriophages direct the synthesis of two proteins holin and lysin. Holin is responsible for making a hole in the cell membrane through which the lysin which is an amide, an endopeptidase or an N-acetylmuramidase can pass through and degrade the cell wall. Thus, application of lysin can lyse the cell wall of uninfected bacterial cells resulting in "lysis from without". This phenomenon is being used in treating and preventing bacterial infections (Loessner *et al.* 2002).

A lysin has a C terminal that has the binding activity and N terminal controls the catalytic activity. So until the C terminal binds with the target, no catalytic activity can take place. The enzyme works by targeting the peptidoglycan cell wall and by making holes in it. Due to high osmotic pressure inside, the holes cause the cell to burst. Another important phenomenon observed in lysis is targeted killing. The lytic enzyme only targets the bacteria for which the phage has specificity (Loessner *et al.* 2002). Research has also shown that bacteria do not become resistant to lytic enzymes even if they are applied repetitively.

Another interesting aspect of phage lytic enzyme is that they don't elicit damaging immune response. A number of experiments show increase in antibody titer against PlyG lysin but the antibodies have no affect on the enzyme and the enzyme works efficiently (Loeffler *et al.* 2002). Studies related to streptococci and pneumococcal lytic enzymes are worth mentioning. Lytic enzyme therapy is also being pursued as a probable anti biowarfare mechanism. Studies have been performed against B. anthracis responsible for anthrax, (Raymond *et al.* 2002). Studies in mouse models have shown promise for post exposure intravenous treatment of anthrax. The lytic enzyme used was PlyG. 90% of mice treated with lytic enzyme survived whereas only 10% of the mice not treated with PlyG survived.

Future of phage therapy in Pakistan

Keeping in mind, the ground realities while working for masses in Pakistan it should be seen that a large population is poverty stricken. We need to work on alternative treatments that are cost effective and economy friendly. Pakistan is basically an agricultural country and our exports include crops and livestock. To increase the yield and decrease the losses associated with crop and livestock, phage therapy can be effectively used. When quality and quantity of our crops and livestock will increase, automatically our foreign exchange reserves will benefit. Phage therapy is cost effective as number of doses needed as compared to antibiotics is far less and time for hospitalizations is also reduced due to speedy recovery. Tuberculosis is a widespread disease of our country and the treatment cost is just preposterous for the masses. Phage therapy against the bacterium can be both affordable and effective for the patients.

Conclusion

Phages being ubiquitous in nature can be easily sought, isolated, identified and put to use. Comparatively easy cell biology of prokaryotes and switching of lytic and lysogenic life cycle of bacteriophages gave many answers about gene regulations and developed a better understanding of molecular biology. Different enzymatic products such as holins and endolysin provide a simple solution for the killing of many Gram positive bacteria.

A lytic phage replicates in a limited time span, producing large population of new phages thus minimizing the number of pathogenic bacteria and controlling the infection. Bacteria although having short generation time still, are no match for the rapid phage replication and in few hours' phages outnumber the pathogenic bacteria. Unless they lurk in inaccessible locations, they are sought out and destroyed.

Unlike antibiotics, bacteriophages are 'living' organisms; they have been infecting bacteria since the beginning of life on this planet. Bacteria evolve to resist phage, but phage evolves too at an amazing rate. Chemists tinkering with new generations of antibiotics can never cope with ever mutating antibiotic resistant bacteria so it seems that phages provide the ultimate antibacterial therapy: lethal, adaptive, highly efficient, safe to humans.

REFERENCES

1. Ahmad SI. 2002. Treatment of post burns bacterial infections by bacteriophages, specifically ubiquitous *Pseudomonas* spp. notoriously resistant to antibiotics. Med Hypotheses; 58:327–31.
2. Aysen Bayram and Iclal Balci 2006: Patterns of antimicrobial resistance in a surgical intensive care unit of a university hospital in Turkey. *BMC Infectious Diseases*, 6:155.
3. Bassett, K. D. 2007. Use of bacteriophage as an antimicrobial in food products. Master's Thesis. Kansas State Univ., Manhattan.
4. Biswas B, Adhya S, Washart P, Paul B, Trostel AN, Powell B: 2002. Bacteriophage therapy rescues mice bacteremic from a clinical isolate of vancomycin resistant *Enterococcus faecium*. *Infect Immun*; 70; 204–10.
5. Cairns J, Stent GS, Watson JD. Phage and the Origins of Molecular Biology. Cold Spring Harbor: *Cold Spring Harbor Laboratory Press*, 1966.
6. Cerveny KE, DePaola A, Duckworth DH, Gulig PA. 2002. Phage therapy of local and systemic disease caused by *Vibrio vulnificus* in iron-dextran-treated mice. *Infect Immun*. 70; 6251–62.
7. Chang S, Sievert DM, Hageman JC, Boulton ML, Tenover FC, Downes FP, *et al.*. 2003. Infection with vancomycin-resistant *Staphylococcus aureus* containing the *vanA* resistance gene. *N Engl J Med*. 348; 1342–7.
8. Chanishvili N, Tediashvili M, Chanishvili T. 2002. Phages and experience for their application in the former Soviet Union. IUMS Congress Paris.
9. Chibani Chennoufi S, Sidoti J, Bruttin A, Kutter E, Sarker S, Brussow H. 2004. In vitro and in vivo bacteriolytic activities of *Escherichia coli* phages: implications for phage therapy. *Antimicrob Agents Chemother*; 48:2558–69.
10. Flaherty, J.E., B.K. Harbaugh, J.B. Jones, G.C. Somodi and L.E. Jackson. 2001. H-mutant bacteriophages as a potential biocontrol of bacterial blight of geranium. *Hort science* 36:98-100.
11. Gill JJ, Svircev AM, Smith R, Castle AJ. 2003. Bacteriophages of *Erwinia amylovora*. *Appl. Environ. Microbiol.* 69:2133–38
12. Ginsburg DS, Calos MP. 2005. Site specific integration with ΦC31 integrase for prolonged expression of therapeutic genes. *Adv Genet*; 54:179-87.
13. Goode, D., V. M. Allen, and P. A. Barrow. 2003. eduction of experimental *Salmonella* and *Campylobacter* contamination of chicken skin by application of lytic bacteriophages. ppl. *Environ. Microbiol*. 69:5032–5036.
14. Groth AC, Fish M, Nusse R, Calos MP. 2004. Construction of transgenic *Drosophila* by using the site-specific integrase from ph ge ΦC31. *Genetics*; 166:1775-82.
15. Groth AC, Calos MP. 2004. Phage integrases: Biology and applications. *J Mol Biol*; 335:667-78.
16. G.M. Eliopoulos, *et al.*1998. Characterization of vancomycin resistant *Enterococcus faecium* isolates from the United States and their susceptibility *in vitro* to dalfopristin quinupristin. Antimicrobial Agents Chemother. 42; 1088–1092.
17. Goode D, Allen VM, Barrow PA. 2003. Reduction of experimental *Salmonella* and *Campylobacter* contamination of chicken skin by application of lytic bacteriophages. *Appl Environ Microbiol*. 69:5032–6.
18. Held PK, Olivares EC, Aguilar CP, Finegold M, Calos MP, Grompe M. 2005. In vivo correction of murine hereditary tyrosinemia type I by fC31 integrase-mediated gene delivery. *Mol Ther*. 11;399-408.
19. Hennes, K.P. and C.A. Suttle. 1995. Direct counts of viruses in natural waters and laboratory cultures by epifluorescence microscopy. *Limnol. Oceanogr*. 40; 1050-1055.
20. Hiramatsu K, Cui L, Kuroda M, Ito T. 2001. The emergence and evolution of methicillin resistant *Staphylococcus aureus*. *Trends Microbiol*. 9;486–93.
21. Ho K. 2001. Bacteriophage therapy for bacterial infections. Rekindling a memory from the pre-antibiotics era. *Perspect Biol Med*. 44; 1–16.
22. J.B. Jones, L.E. Jackson, B. Balogh, A. Obradovic, F.B. Iriarte, and M.T. Momol. 2007. Bacteriophages for Plant Disease Control. *Annu. Rev. Phytopathol*. 45:245–62
23. Kacica M. Vancomycin resistant *Staphylococcus aureus* New York, 2004. MMWR (Morb Mortal Wkly Rep). 53:322–3.
24. Leverentz, B., W.S. Conway, M.J. Camp,W.J. Janisiewicz, T. Abuladze, M.Yang, R. Saftner and A.Sulakvelidze.2003. Biocontrol of *Listeria monocytogenes* on fresh cut produce by treatment with lytic bacteriophages and a bacteriocin. *Appl. Environ. Microbiol*. 69: 4519-4526.

25. Leverentz B, Conway WS, Janisiewicz W, Camp MJ. 2004. Optimizing concentration and timing of a phage spray application to reduce *Listeria monocytogenes* on honeydew melon tissue. *J Food Prot*. 67:1682–6.
26. Loessner MJ, Kramer K, Ebel F, Scherer S. 2002. C-terminal domains of Listeria monocytogenes bacteriophage murein hydrolases determine specific recognition and high affinity binding to bacterial cell wall carbohydrates. *Mol Microbiol.*44:335–349
27. Loeffler JM, Djurkovic S, Fischetti VA. 2003. Phage lytic enzyme Cpl-1 as a novel antimicrobial for pneumococcal bacteremia. *Infect Immun.* 71:6199–6204.
28. Merril, C.R., D. Scholl and S.L. Adhya.2003. The prospect for bacteriophage therapy in Western Medicine. Nat.Rev. 2: 489-497. Wagner, P.L. and M.K. Waldor. 2002. Bacteriophage control of bacterial virulence. *Infect. Immun*. 70: 3985-3993.
29. Merril CR, Biswas B, Carlton R, Jensen NC, Creed GJ, Zullo S, 1996. Long-circulating bacteriophage as antibacterial agents. Proc Natl Acad Sci USA. *Biol. Rev*. 93:3188–92.
30. M Pai, SP Kalantri 2005. Bacteriophage-based Tests for Tuberculosis. *Indian Journal of Medical Microbiology*. 23: 3: 149-150.
31. Nakai T, Sugimoto R, Park KH, Matsuoka S, Mori K, Nishioka T, 1999. Protective effects of bacteriophage on experimental *Lactococcus garvieae* infection in yellowtail. *Dis Aquat Org*. 37:33–41.
32. Park SC, Nakai T. 2003. Bacteriophage control of *Pseudomonas plecoglossicida* infection in ayu *Plecoglossus altivelis*. *Dis Aquat Org*. 53:33–9.
33. Payne RJ, Phil D, Jansen VA. 2000. Phage therapy: the peculiar kinetics of self-replicating pharmaceuticals. *Clin Pharmacol Ther*. 68:225–30.
34. Payne RJ, Jansen VA. 2001. Understanding bacteriophage therapy as a density dependent kinetic process. *J Theor Biol.* 208:37–48.
35. Payne RJ, Jansen VA. 2002. Evidence for a phage proliferation threshold. *J Virol.* 76:13123–4.
36. Payne RJ, Jansen VA. 2003. Pharmacokinetic principles of bacteriophage therapy. *Clin Pharmacokinet*. 42:315–25.
37. Pillai SK, Sakoulas G, Wennersten C, Eliopoulos GM, Moellering RC Jr, Ferraro MJ, 2002. Linezolid resistance in *Staphylococcus aureus*: characterization and stability of resistant phenotype. *J Infect Dis*. 186:1603–7.
38. Raymond Schuch, Daniel Nelson & Vincent A. Fischetti: 2002. A bacteriolytic agent that detects and kills *Bacillus anthracis.* Letters to *Nature.* 418. 884-889.
39. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE. 2003. Restriction enzymes and methyltransferases. *Nucleic Acids Res*. 31:418-20.
40. Sharp, R.J. 2001. Bacteriophages: biology and history. *J. Chem.Technol. Biotechnol*. 76: 667-672.
41. Shigenobu Matsuzaki, Mohammad Rashel, Jumpei Uchiyama, Shingo Sakurai, Takako Ujihara, Masayuki Kuroda · Masahiko Ikeuchi · Toshikazu Tani Mikiya Fujieda, Hiroshi Wakiguchi, Shosuke Imai: 2005. Bacteriophage therapy: a revitalized therapy against bacterial infectious diseases. *J Infect Chemother*. 11:211–219
42. Shimada K, Nakano K, Igari J, Oguri T, Ikemoto H, Mori T, etal. 2004. Susceptibilities of bacteria isolated from patients with lower respiratory infectious diseases to antibiotics. *Jpn J Antibiot*. 57:213–45.
43. Sklar, I. A. N. B., and R. D. Joerger. 2001. Attempts to utilize bacteriophage to combat *Salmonella enterica* serovar Enteritidis infection in chickens. *J. Food Saf*. 21:15–29.
44. Soothill JS. 1992. Treatment of experimental infections of mice with bacteriophages. *J Med Microbiol*. 37:258–61.
45. S. Tslodras *et al.* 2008. Linezolid resistance in a clinical isolate of *Staphylococcus aureus*. *Lancet*. 358; 207–208.
46. Ward Thomas W., Bruhn Debby F, and Bulmer Deborah K. July 21, 1992.
Patent: Lysogenic bacteriophage isolated from acidophilium.
47. Westwater, C., L.M. Kasman, D.A.Schofield, P.A. Werner, J.W. Dolan, M.G. Schmidt and J.S. Norris. 2003. Use of genetically engineered phage to deliver antimicrobial agents to bacteria: An alternative therapy for treatment of bacterial infections. Antimicrob. *Agents Chemother*. 47:1301-1307.
48. Wills QF, Kerrigan C, Soothill JS. 2005. Experimental bacteriophage protection against *Staphylococcus aureus* abscesses in a rabbit model. *Antimicrob Agents Chemother*. 49:1220–1.
49. Zilka N, Vechterova L, Kontsekova E, Novak M: 2003. A rapid immunohistochemical primary screening assay for hybridomas. *J. Immunol Methods*. 272:49-53.

Antimicrobial	MRSA n = 8 2	MSSA n = 1 8
Vancomycin	0%	0%
Clindamycin	73.2%	66.7%
Erythromycin	90.2%	66.7%
TMP/SMX[3]	42.7%	22.2%
Ciprofloxacin	79.3%	72.2%
Tetracycline	87.8%	77.7%

Fig. 1. MRSA=Methicillin-resistant S. aureus, MSSA= Methicillin-susceptible S. aureus, P/SMX=Trimethoprim/sulfamethoxazol. Antimicrobial resistances of MRSA and MSSA isolated from infections in ICU (Aysen et al. 2006).

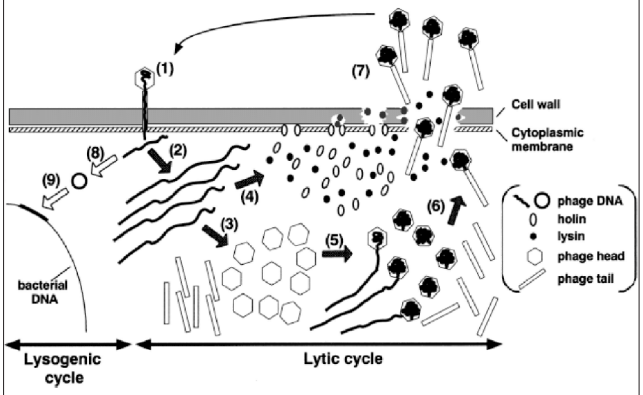

Fig.2. Schematic illustration of phage induced bacteriolysis. (1) Adsorption and DNA injection; (2) DNA replication; (3) production of head and tail; (4) synthesis of holin and lysin; (5) DNA packaging; (6) completion of phage particle; (7) disruption of the cell wall and release of the progeny; (8) circularization of phage DNA; (9) integration of the phage DNA into the host genome. (Shigenobu et al.2005).

Pollen allergy; Mechanism and Etiology

Shazia Choudhary, Mamoona Noreen, Muhammad Arshad
NUST Center of Virology and Immunology (NCVI)
National University of Sciences and Technology Islamabad, Pakistan
shazia_126@hotmail.com, mamoona786@hotmail.com, arshad_pk2002@yahoo.com

Abstract

The term allergy refers to an exaggerated immune response to a foreign molecule. In a normal healthy individual, this foreign molecule will act as a harmless antigen, as it will be recognized & cleared by the immune system without causing any significant damage to host tissues. In certain individuals, such antigens stimulate immune system in such a way that a series of exaggerated immune responses are generated against it, leading to significant tissue injury and damage to the host. Such an exaggerated immune response may also cause death of the patient. Such antigen is termed as an allergen. Allergy is categorized under immunoglobulin E (IgE) mediated type I hypersensitivity reactions. These reactions occur when an antigen acting as an allergen attacks the immune system of the host and causes excessive stimulation of mast cells and basophils.

This is followed by the release of allergy mediators which are responsible for causing local or systemic anaphylaxis, allergic asthma, allergic rhinitis, conjunctivitis.Type1 allergens are of various types, the mechanism of action is same; however underlying factors may differ in each type. This article is about pollen allergens, in specific & will discuss the common sources of pollen allergens in Pakistan, the prevalence and the factors behind pollen allergy.

Key words: Allergy; allergen; aeroallergen; IgE antibody; Pollen grain; mast cell.

Mechanism of action and types of class 1 allergens

The actual mechanism consists of two different phases. The first phase is named as the sensitization phase and consists of events associated with the first time entry of allergen in the body. During this phase the allergen stimulates B lymphocytes (B cells) cells to produce antibody secreting plasma cells, which ultimately produce specific IgE antibodies. These antibodies get attached on the surface of mast cells and basophills. Upon subsequent exposure the allergen gets attached to the IgE molecules already bound on the surface of mast cells and results in cross linking. This is followed by a series of events which ultimately lead to the release of allergic mediators which may produce systemic anaphylaxis, localized anaphylaxis, allergic rhinitis, allergic asthma, atopic dermatitis, etc (Mauro et al. 2000; Mike 2006). Mechanism of action of Type 1 allergens is given in Figure 1.Type I allergens can be subdivided into different categories depending upon the route of entry (Figure. 2). This article will focus only on the pollen grains.

Pollen grain as potential source of allergy

Pollen grains are one of the most commonly encountered aeroallergens. As the name indicates, aeroallergens are airborne allergens, which enter the body through respiratory route. How far they travel in respiratory tract, is dependent on their size. The smaller the size of an aeroallergen, faster it will travel through the respiratory tract. The particles with size < 8 µm are considered to be able to penetrate in the respiratory tract, whereas those with size ≥ 20 µm are deposited on naso or oropharynx or on ocular mucous membrane, leading to allergic rhinitis, allergic asthma or conjunctivitis (Eric *et al*. 2004).

Pollens are one of the most abundantly found particles

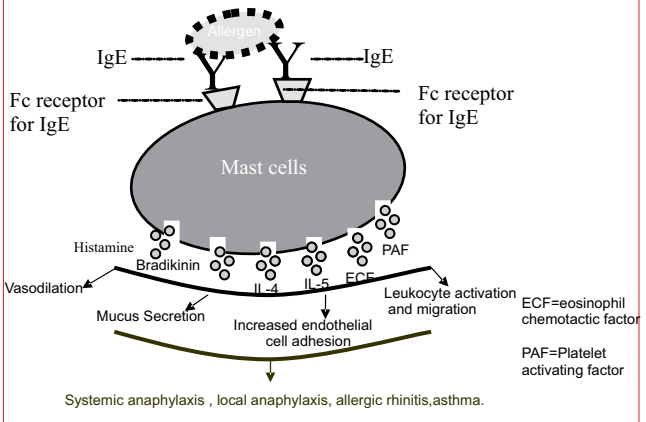

Fig.1. Mechanism of action of Type I allergens

in air. Being extremely light, they can travel up to large distances in air. Although pollen grains are potential allergens but all of the pollen grains are not causing allergy. Similarly allergy causing pollens will not affect all of the individuals. Allergic disorders develop only in case of heavy exposure to the allergens or in case if the individual has some genetic predisposition to allergy (Burge 2001).

Sources of allergy causing pollens

Allergy causing pollen can come from any plant, grasses, trees, weeds, flowering plants etc. The most common sources of allergy causing pollens in Pakistan are given in Table 1. (Shahid 2010).

Prevalence of allergy causing pollens

Pollination time of different plants varies largely. Some of the plants pollinate in the spring season whereas others pollinate in the fall. Therefore, pollen count of a particular plant keeps on changing throughout the year. The

Corresponding Author: Shazia Choudhary

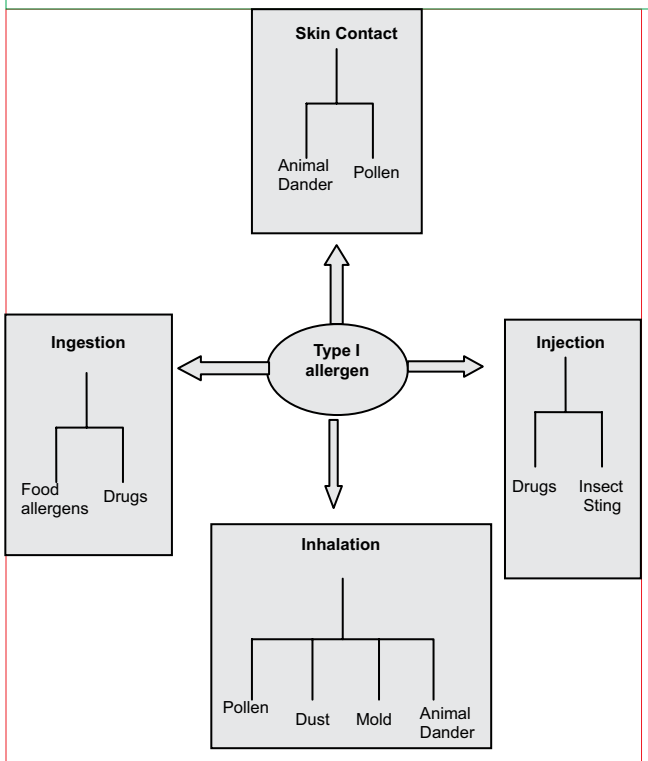

Fig .2. Different types of Type 1 allergens and their possible mode of entry in host.

pollination time of most common allergy causing plants in Pakistan is given in Table 2 (Shahid 2010).

During pollination, the pollen count in air is too high, indicating an alarming situation for the allergy patients. Different types of air samplers are used to collect the pollens present in air which are then counted and identified in microscope.

Consequences of pollen allergy

The chances of development allergy related disorders increase in season when pollen count is high. In case of allergic rhinitis, the symptoms may include sneezing, rhinorrhea, congestion etc (Bousquet *et al*. 2001). In case of asthma, patients experience episodic symptoms of wheezing, cough and dyspnea (CDC guidelines 2003). Sometimes patients experience acute inflammation of eyes and frequent itching, a condition termed as conjunctivitis (Rose 2007).

Since pollens travel through air, it is not possible to avoid 100% exposure. However some precautions may really help the patients in staying safe, or at least in lowering the severity of the symptoms, e.g., wearing a mask before going out, drinking a lot of water to prevent dehydration, using anti allergy medicines and inhalers, etc.

Factors influencing the development of allergy

In order to have a better understanding of pollen allergy it is important to mention that multiple factors are involved in its pathogenesis and susceptibility to individuals. These factors may include the following.

Genetic factors

Some people are genetically predisposed to several types of allergies including pollen allergy. They have higher levels of circulating IgE (up to12 µg/ml) than healthy individuals (0.3 µg/ml). The immune system of allergic individuals is tilted toward Th2 subtype. This results in the production of interleukin-4 (IL-4) & IL-13 which promote class switching in the B cells and thus synthesizing IgE antibodies.

Several studies have also reported role of single nucleotide polymorphisms of these cytokines in the pollen allergy, e.g., IL-13 polymorphism (C112T) in IL-13 has been found to be a protective factor in olive allergy whereas IL-13 (R130Q) is found to be a risk factor (Elena *et al*. 2009). Similarly IL-13 (Arg130Gln) single nucleotide polymorphism has also been linked to the increased risk of allergy (Min Wang *et al*. 2003). Apart from cytokines, several studies have also reported the role of mutations in Toll like receptors (TLR) in allergy, e.g., it has been reported that TLR-2 (R753Q) mutation modifies the cytokine production and increases the risk of atopic dermatitis (Salima 2008).

Hygiene hypothesis

It is strongly believed that lack of early childhood exposure to infectious agents or in other words individuals who are brought up in extremely hygienic environment are more prone to develop different types of allergies including pollen allergy (Maria *et al*. 2002). It is argued that lack of exposure to bacterial/viral pathogens down regulates the T-helper-1 (Th1) arm of immune system while T-helper-2 (Th2) arm is up regulated. Hence the individuals become more prone to allergies (Sergio *et al*. 2004).

Air polution

Diesel and other exhaust have been shown to enhance the ability to make the allergy antibody, IgE, in response to exposure to allergens (Syed Zafar *et al*.2009; Bartra *et al* 2007). It has also been reported that diesel exhaust particles influence the expression of huge amount of allergy causing proteins in pollen grains. Air pollutants also cause mucosal damage and impaired mucociliary clearance ultimately influencing increased risk of allergy (Riccardo 2007).

Conclusion

Pollen allergy is one of the most commonly encountered problems not only in Pakistan, but also in the rest of the World. Efforts are required to identify allergy causing plants (other then those which are already known) so that their plantation can be controlled. Full characterization of pollen allergens is desired so as to develop a better understanding of pathogenesis of pollen allergy. This may help in designing better treatment strategies and preventive measures.

Abbreviations

1. Arginine R
2. Arginine Arg
3. B lymphocytes (B-cell)

Table 1. Sources of Pollen Allergy in Pakistan

GRASSES	Tree	Weeds
Bermuda grass	Cyprus	Rose
Italian grass	Populus	Rumex
Flavus grass	Olia Ferogenia(Olive)	Urticacae(stinging nettle)
Wild oat	Jackarande	Ricinum (Castor)
Trimothy Grass	Silver Oak	Chenopod(Bathoo)
Meadow Grass	Accer	Artemesia
Italian Rye Grass	Paper Mulbery	Lantana
Bromus Grass	Accacia	Amaranthus viridus
SorgumGrass	Accacia Modeta	Brassica Compestoris(Sarson)
Swamp Meadow Grass	Pine	Dandilion
Smooth Finger Grass	Eucalyptus	Plantain
Common Reed Grass	Chinese Tallow	Canabis(bhang)
Wild Cane	-------	---------

Table 2. Pollination time of most common allergy causing plants in Pakistan.

Grasses	Bromus, Italian Grass, Flavus, Wild Oat, Trimothy Grass, Meadow Grass	Spring
	Sorgum, Swamp Meadow Grass, Common Reed, Wild Cane, Lemon Grass	Fall
	Bermuda	Throughout the year
Trees	Paper Mulberry, Pine, Acacia, Eucalyptus, Chinese Tallow, Olive, Jackarande	Spring
	Rubinia, Accacia Modesta, Silver Oak	Fall
Weeds	Rose, Brassica, Rumex, Dandilon, Urcticaceae, Plantain, Ricinum	Spring
	Chenopod, Cannabis, Amaranthus.	Fall

4. Glutamine (Q)
5. Glutamine (Gln)
6. Immunolglobulin E (IgE)
7. Interleukin-4 (IL-4)
8. Interleukin-13 (IL-13)
9. T-Helper-1 (Th-1)
10. Toll like receptor (TLR)

REFERENCES

1. Mauro Cresti and Hans Ferdinand Linskens. 2000. Pollen-allergy as an ecological phenomenan: a review. Plant Biosystems, 134:341-352.
2. Thomas M. 2006. Allergic rhinitis: evidence for impact on asthma. BMC Pulm Med, 6 Suppl 1: S4.
3. M. Eric Gershwin, Stanley M. and Naguwa. 2004. Allergy and Immunology secrets, 13-100pp.
4. Burge HA. 2001. An update on pollen and fungal spore aerobiology. J Allergy Clin Immunol, 110:544-552.
5. Abbas, Allergies and allergic reactions, http://allergycenter.info/doctor_pollen.asp. Allergy and Asthma Clinic, Pakistan, Accessed 2 January 2010.
6. Bousquet J,van Couwenberge P and Khaltaev N. 2001. Rhinitis and its impact on asthma. J Allergy Clin Immunol, 108:S147-336.
7. Centers for disease control and prevention: Key clinical activities for quality asthma care: Recommendation of National Asthma Education and Prevention Programme MMWR52 (RR-6):1-8, 2003.
8. Rose P 2007. Management strategies for acute infective conjunctivitis in primary care: a systematic review. Expert Opin Pharmacother, 8:1903–21.
9. 11- Llanes E, Quiralte J, Lopez E, Sastre B, Chacartegui M, del Pozo V, Palomino P, Lahoz C and Cardaba B. 2009. Analysis of Polymorphisms in Olive Pollen Allergy: IL13, IL4RA, IL5 and ADRB2 Genes. Int Arch Allergy Immunol, 148:228-38.
10. 12- Wang M, Xing ZM, Lu C, Ma YX, Yu DL, Yan Z, Wang SW and Yu LS. 2003. A common IL-13 Arg130Gln single nucleotide polymorphism among Chinese atopy patients with allergic rhinitis. Hum Genet, 113:387-90.
11. 13- Mrabet-Dahbi S, Dalpke AH, Niebuhr M, Frey M,

Draing C, Brand S, Heeg K, Werfel T and Renz H. 2008. The Toll-like receptor 2 R753Q mutation modifies cytokine production and Toll-like receptor expression in atopic dermatitis. J Allergy Clin Immunol, 121:1013-9.

12. 14- Yazdanbakhsh M, Kremsner PG and van Ree R 2002. Allergy, Parasites and the Hygiene Hypothesis. Science, 296:490-4.
13. 15- Romagnani S. 2004. The increased prevalence of allergy and the hygiene hypothesis: missing immune deviation, reduced immune suppression, or both? Immunology, 112:352-63.
14. 17- Zafar SI, Iqbal AK, SM Nasir, Qurashi T and Durrani R. 2009. Air pollution assessment in urban areas and its impact on human health in the city of Quetta, Pakistan. Clean Technologies and Environmental Policy.
15. 18- Bartra J, Mullol J, del Cuvillo A, Davila I, Ferrer M, Jauregui I, Montoro J, Sastre J and Valero A. 2007. Air pollution and allergens. JIACI, 17 Suppl 2:3-8.
16. 19- Polosa.R 2007. The interaction between particulate air pollution and allergens in enhancing allergic and airway responses. Curr Allergy Asthma Rep, 1:102-7.

Cloning and Overexpression of Pyrococcus Furiosus Endoglucanase A Gene (eglA) in Escherichia Coli

Najam-us-Sahar, Sadaf Zaidi*[1], M. Waheed Akhtar[2]
[1]NUST Centre of Virology and Immunology
National University of Science and Technology, Islamabad
[2]School of Biological Sciences, University of the Punjab, Lahore
zaidi.sahar@gmail.com, mwa@brain.net.pk

Abstract

This study describes the cloning and high-level expression of an endoglucanase A gene (eglA) from a hyperthermophilic archeon Pyrococcus furiosus. An expression plasmid pET-EglA was constructed for the production of recombinant EglA in E. coli BL21 (DE3) under the control of T7lac promoter. Following induction, ~35kDa protein expressed at levels greater than 20% of the total E. coli cellular proteins. The expressed protein, however, was in the form of inclusion bodies with little enzymatic activity, which was solubilized using higher concentration of denaturing agent (8M urea) followed by its refolding to an active state. A 7-8 fold increase in enzyme activity corresponding to 285U/mg specific activity could be achieved after refolding. The refolded EglA, partially purified by heat treatment upto ~92%, is being investigated for applications like hydrolysis of cellulose, a major component of plant biomass. Local, upscale and cheap production of these cellulolytic enzymes can help in reducing the costs of many processes in various industries like poultry and textile.

Key words: Endoglucanase A, Pyrococcus furiosus, overexpression, inclusion bodies, refolding

Introduction

Pyrococcus furiosus is an obligately anaerobic, heterotropic hyperthermophilic archeon that grows optimally at 98-100°C and employs a fermentative type of metabolism utilizing a range of sugars such as cellobiose, laminarin, chitin, maltose, barley glucan and starch as primary carbon and energy sources [1-5]. Various enzymes like *Pfu* DNA polymerases, amylases, laminarases, α-, β-glucosidases, etc. have been isolated from this marine archaebacteria, till to date. These enzymes have attracted considerable research and commercial interest because of their evolutionary significance and potential applications in the biotechnology industry [6,7].

Cellulose, a major component of plant biomass and the most abundant polysaccharide on biosphere [8], is composed of D-glucose units that are linked together via β-1,4-glucosidic linkages to form linear chains. Earlier it was documented that *P. furiosus* is capable of degrading α-linked glucose polymers, such as starch and glycogen, only by the concerted action of alpha glucosidase, α-amylase and pullulanase [9]. But later it was found that *P. furiosus* is also able to grow on β-linked glucose polymer and has ability to hydrolyze laminarin to glucose by a combined action of extracellular endo-β-1,3-glucanase (LamA) and β-glucosidase (CelB). Some other researchers have identified a novel endoglucanase (EglA), capable of degrading β-1, 4 linkages of both cellulose and mixed-linkage β-glucans, in *P. furiosus* [10].

The two endoglucanases, LamA and EglA, have distinctly different amino acid sequences and substrate specificities and therefore, belong to two different enzyme families, i.e., family 16 and 12, respectively [11]. The nucleotide sequence of the EglA encodes a 319 amino acid long protein with a calculated molecular mass of 35.9 kDa and an N-terminal signal peptide of 19 amino acids. Cellulose binding domain, however, is absent in this enzyme.

In the present study, the gene encoding EglA from *P. furiosus* (ATCC 43587) was cloned and expressed in *E. coli* BL-21 (DE3) under the regulation of T7*lac* promoter. Refolding and partial purification of recombinant enzyme has also been described.

Materials and methods

Bacterial strains, DNA, enzymes and kits

E. coli strains DH5α and BL21 (DE3) were used as cloning and expression hosts, respectively. The genomic DNA of *P. furiosus* was obtained from the American Type Culture Collection (ATCC 43587). *Taq* DNA polymerase, T4 DNA ligase and restriction endonucleases were purchased from MBI Fermentas (MD, USA). QIAquick gel extraction and QIAprep miniprep kits, used for DNA extraction and plasmid miniprerparation, respectively were acquired from QIAGEN Inc. (CA, USA). All other chemicals used were of the analytical grade commercially available.

Construction of recombinant pET-EglA plasmid

T7*lac* promoter based pET-22b(+) vector was used for the construction of pET-EglA expression plasmid. A pair of forward (NSSZ-F: 5´GCAG*CATATG*ATATATTTTGTAGAAAAGTATCAT ACCTC-3´) and reverse (NSSZ-R: 5´-GA*GGATCC*TTGATAATTCCCAGGTTGCCGA-3´) primers was designed to amplify the *egl*A gene (GenBank Accession No. AF181032) from the genomic DNA of *P. furiosus*. *Nde*I and *Bam*HI sites were incorporated at 5´-

Corresponding Author: Najam-us-Sahar

termini of the designed primers (shown in italics) to facilitate the directional cloning. The conditions used for PCR were: 95°C for 4 minutes, 30 cycles of [94°C 45 sec., 60°C 45 sec., and 72°C 1 min.], and a final extension of 72°C for 20 min. The amplicon thus obtained was cloned in pTZ57R/T vector by employing dA.dT tailing technique to generate pTZ-EglA, followed by its transformation into *E. coli* DH5α competent cells according to the standard molecular biology protocols [12].

For the construction of pET-EglA, the pTZ-EglA plasmid was digested with *Nde*I and *Bam*HI restriction enzymes to release the *egl*A, which was then inserted into a similarly digested pET-22b(+) vector by ligase-mediated cloning. The resulting recombinant plasmid (pET-EglA) was used to transform *E. coli* BL21 (DE3) expression host and the transformants were selected on LB-agar plates containing 100µg/ml ampicillin. Presence of insert in positive transformants was confirmed either by colony PCR, restriction digestion or nucleotide sequencing using Beckman CEQ 8000 Genetic Analyzer.

Analysis of EglA expression

E. coli BL21 (DE3) harboring the pET-EglA were grown in 10ml LB-ampicillin medium overnight at 37°C in an orbital incubator shaker (150 rpm). The following day, 3ml of this overnight culture was used to inoculate 100ml LB-ampicillin broth, grown at 37°C until the OD_{600} reached 0.5-0.8. At this stage, the cells were either induced with 5-10mM lactose or 0.5mM IPTG and the fermentation was continued with constant shaking at 150rpm till the maximum cell density was achieved. The cells were thereafter harvested by centrifugation (6,500 rpm, 4°C, and 15 minutes), resuspended in 50mM Tris-Cl (pH 7.0) and subjected to sonication (15x30s bursts with 1 minute interval between the two successive pulses). The soluble and insoluble fractions were separated at this point and analyzed by 15% SDS-PAGE [13]. The inclusion bodies containing the recombinant EglA were washed twice with Triton X-100 and then solubilized using higher concentration of urea (8M).

Refolding and partial purification of EglA

The solubilized EglA was refolded in a renaturation buffer containing 50mM Tris-Cl (pH 9.0), 2mM EDTA, 5mM cysteine, 0.5mM cystine and 1mM PMSF. Solubilized protein solution was added to the renaturation buffer in pulses (1ml/hr, 4°C) with constant stirring on a magnetic stir plate. Following refolding, the sample was dialyzed against Tris-Cl (pH 7.0), concentrated in a freeze dryer and then subjected to heat treatment at 70°C for different time intervals ranging between 10-60 min., for partial purification of recombinant enzyme. Protein estimation and enzyme activity assessment was performed at each step by Bradford [14] and DNS methods [15], respectively.

Zymogram analysis

For zymogram analysis, 15% SDS-gel, after electrophoresis, was sequentially transferred to solution A [50mM Tris-Cl (pH 7.0) containing 20% isopropanol] and solution B [50mM Tris-Cl (pH 7.0)] for one hour each, at 4°C in a petridish. The SDS-gel was then placed on an agar-CMC gel containing 1.7% agar and 0.5% carboxy methyl cellulose (CMC) and incubated at 60°C for 1 hour. Following incubation, the SDS-gel was removed and the agar-CMC gel was stained with Congo red solution for 10 minutes, and destained with 1M NaCl.

Results and discussion

PCR amplification and cloning of eglA gene

The gene encoding full-length EglA protein (without signal peptide) was amplified from *Pyrococcus furiosus* genomic DNA using a pair of gene specific primers. PCR amplification yielded a single product of around 0.90 kb (Figure. 1). To obtain the maximum amplification of *egl*A gene, the PCR reaction conditions were optimized with respect to annealing temperatures and $MgCl_2$ concentrations and the best amplification could be achieved at an annealing temperature of 60°C using 2.5 mM $MgCl_2$.

The amplicon was gel purified, cloned in pTZ57R/T vector and then subcloned between *Nde*I and *Bam*HI sites of pET-22b(+) vector to generate pET-EglA expression plasmid (Figure 2). The recombinant plasmid was first maintained in *E. coli* DH5α for vector propagation and then transformed into BL21 (DE3) for expression studies.

The transformants were selected on LB-ampicillin plates and the presence of *eglA* gene was confirmed by restriction digestion and/or colony PCR. Five different colonies were screened for the presence of insert by colony PCR and the results are shown as Figure 3. A prominent band of 0.90kb in all the screened colonies confirmed the successful transformation.

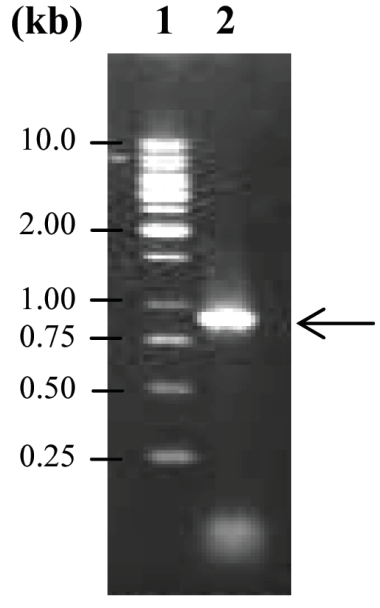

Fig. 1. PCR product of eglA gene analyzed on 1% agarose gel. Lane 1, 1kb DNA ladder; lane 2, ~0.90kb PCR amplified product. Arrow indicates the position of amplicon.

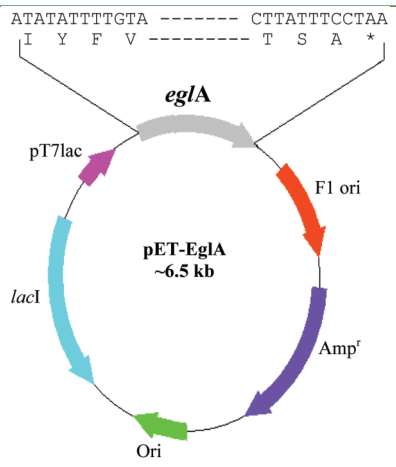

Fig. 2. Construction of pET-EglA recombinant plasmid. 0.90 kb amplicon was cloned in pET-22b(+) plasmid containing origin of replication (ori), T7-lac promoter (pT7lac), ampicillin resistance gene (Ampr) and gene for LacZ (lacI).

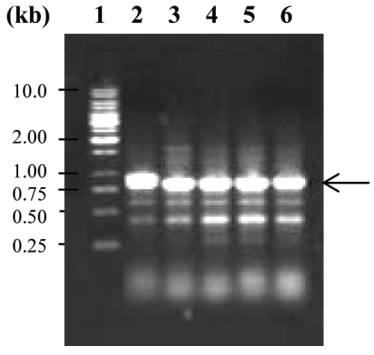

Fig. 3. Agarose gel showing colony PCR results of eglA gene after cloning in pTZ57R/T vector. Lane 1, 1kb DNA ladder; lanes 2-6, eglA gene from different colonies after transformation into E. coli competent cells. Arrow indicates the position of amplified PCR product.

Expression analysis

E. coli BL-21 (DE3) cells transformed with pET-EglA plasmid were grown in LB-ampicillin broth till an OD_{600} of 0.5-0.8 and induced with IPTG or lactose to analyze the expression of recombinant EglA in transformed *E. coli*. Sample fractions were collected at different time intervals during the fermentation and subjected to 15% SDS-PAGE.

Analysis of total cell protein revealed the presence of ~35 kDa protein (expected molecular mass of EglA) in both lactose- and IPTG-induced cells (Figure 4). A band of similar size, however, was absent in control and uninduced cells. It was further observed that the expression of recombinant EglA was better in lactose-induced cells where it accounted for ~25% of the total *E. coli* cellular proteins as compared to ~20%, when induced with IPTG (compare lane 1 of Figure 4 A and B). Owing to the expensiveness of IPTG as inducer and also to the effectiveness of lactose in achieving high level of expression, lactose was used in all subsequent expression studies.

SDS-PAGE analysis of soluble and insoluble fractions revealed that the recombinant EglA was expressed in the form of inclusion bodies (IBs), as a band of 35kDa corresponding to EglA protein was absent in the fraction obtained after centrifugation of the sonicated sample (Fig. 4B, lanes 2 and 3). This probably was the result of high level expression of recombinant EglA under the T7lac promoter providing not enough time for the folding of expressed protein in *E. coli*. Since expression of EglA was undetectable in soluble fraction, we proceeded with IBs for subsequent purification and refolding of expressed enzyme.

Purification and refolding

The IBs obtained from the lysate of cells grown in 2liter LB-ampicillin medium were washed twice with Triton X-100 to obtain relatively pure IBs. The purification achieved at this stage was around 80% (Figure 5). Washed IBs were thereafter solubilized in 8M urea and subjected to refolding in a renaturation buffer as detailed in materials and methods.

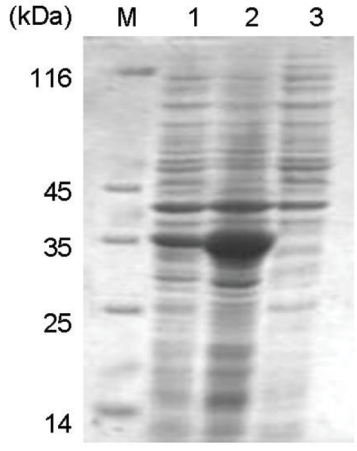

Fig. 4. 15% SDS-gel electrophoresis of total E. coli cellular proteins showing the expression of recombinant EglA. Lane M, molecular weight marker; lane 1, total cell protein; Lane 3, insoluble fraction; Lane 4, soluble fraction.

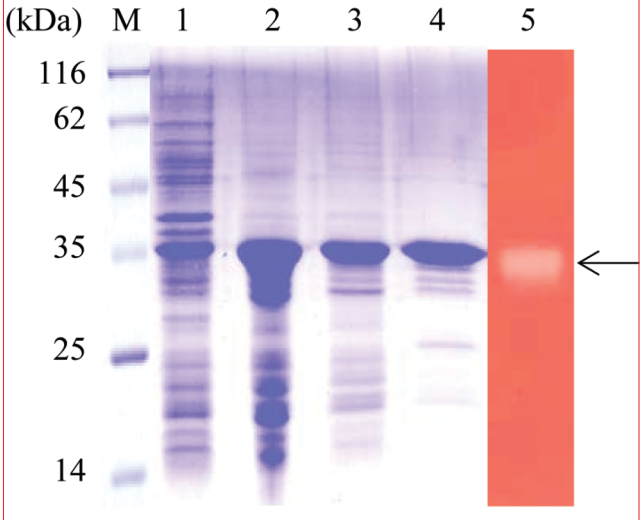

Fig. 5. SDS gel electrophoresis of recombinant EglA at different stages of processing. Lane M, protein size markers; lane 1, total E. coli cell proteins; lane 2: inclusion bodies (IBs); lane 3: IBs after washing with Triton X-100; lane 4: recombinant EglA after refolding; lane 5, zymogarm analysis of refolded protein.

Table 1. Comparison of specific activities of recombinant EglA of *Pyrococcus furiosus* before and after refolding.

Sample type	Volume	EglA (% of total protein)	Enzyme activity (U/ml)	Protein concentration (mg/ml)	Specific activity (U/mg)
Total cell protein	2 liter	17	-	-	-
Sonicated sample (insoluble fraction)	35 ml	65	1.66	0.045	
Insoluble fraction after washing with Triton X 100	35 ml	80	1.70	0.043	
Refolded sample (after dialysis)	75 ml	85	6.2	0.026	
Heat treated sample	75ml	~92	6.8	0.024	283.33

The addition of IBs in a pulsatile manner and the presence of cysteine and cystine in the renatuartion buffer helped in proper refolding of EglA while preventing the enzyme from precipitation/aggregation.

Zymogram analysis revealed that following refolding, the recombinant EglA was not only active but also sufficiently pure (Fig. 5, lanes 4 and 5). Heat treatment of refolded sample at 70°C for 15-60 min. further improved its purification without activity loss.

Activity assay

The specific activity of recombinant EglA was estimated by DNS method at each stage of purification and refolding. The results obtained are summarized in Table 1. The data reveals that the specific activity enhanced almost 8 fold after proper folding of the recombinant EglA.

Conclusion

Locally produced cellulase enzymes have the advantage of being cheap and readily available for use at various levels in industrial processes. The current study describes high level expression of a cellulase enzyme from a hyper thermophilic organism, which has the added benefit of use in industrial applications where heat generation has been reported to inactivate the enzymes being used.

Acknowledgement

This work was supported by a research grant from the Higher Education Commission, Government of Pakistan.

REFERENCES

1. Fiala G and K.O. Stetter, (1986), *Pyrococcus furiosus* sp. Nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100^0 *C. Arch. Microbiol*. 145: 56-61.
2. Claire Vieille and G. J. Zeikus, (2001) Hyperthermophilic enzymes; sources, uses, and molecular mechanisms for thermostabililty, *Microbiol Mol Bio Reviews*, p.1-43.
3. Koning, S.M., W.N. Konings, and A.J.Driessen (2002) Biochemical evidence for the presence of two alpha-glucoside ABC-transport systems in the hyperthermophilic archaeon *Pyrococcus furiosus*. *Archaea* 1: 19-25.
4. Costantino H.R., Stephen H.B. and M. K Robert. (1990) Purification and charactrerization of a glucosidase from a hyperthermophilic archaebacterium *Pyrococcus furiosus*, exhibiting a temperature optimum of 105 to 115C. *J. Bacteriol*, p.3654-3660.
5. Ghose T.K. Measurement of cellulose activities (1987), *Pure Appl. Chem*. 59, 257/
6. Bradford, M.M. (1976) *Anal. Biochem*. 72, 284.
7. Michael W. Bauer, L. B.Driskill, W. Callen, M. A. Snead, E J. Mathur and R. M. Kelly. 1999. An Endoglucanase, EglA, from the hyperthermophilic Archaeon *Pyrococcus furiosus* hydrolyzes B-1,4-bonds in mixed linkage (1- 3), (1- 4) B-D-glucans and *Cellulose J. Bactreriol*, p 284-290 vol. 181 No.1.
8. Sambrook, J. and D. W. Russel (2001) Molecular cloning: A Laboratory manual, 3rd edition CSHL press (NY

Molecular Genetics of Rheumatoid Arthritis and Future Prospectus in Pakistan

Peter John*, Altaf Ahmad*, Mushtaq Ahmad, Syed Fazal Jalil*, Rukhsana Kausar*, Zeeshan Gauhar***
Javaid Mehmood Malik*, Moghees Ikram Ameen****, Ishtiaq Qadri*, Attya Bhatti***
*NUST Center of Virology & Immunology (NCVI)
National University of Sciences & Technology, Islamabad Pakistan
**Department of Rheumatology, Military Hospital Rawalpindi
***Department of Rheumatology & Rehabilitation Foundation University Medical College, Rawalpindi
****KRL Hospital, Islamabad
peter-ncvi@nust.edu.pk

Abstract

The spectrum of rheumatic disease is wide and includes conditions with diverse pathology, although most have in common a heritable risk with a complex genetic basis. Over the past decade intense efforts has been done to understand the contribution of genotype to the expression of disease in terms of both basic pathogenesis and clinical characteristics. The dramatic improvement in technology and methodology has accelerated the pace of gene discovery in complex disorders in an exponential fashion. This review focuses on rheumatoid arthritis and describes some of the recently described genes that underlie this condition and the extent to which they overlap.

Rheumatoid arthritis

Rheumatoid Arthritis (RA) is a multifactorial disease due to a combination of genetic and environmental factors (Dieude and Cornelis, 2005). Its genetic component has been suggested by familial aggregation, twin studies, and segregation analysis (Cornelis et al. 1998). It has been characterized by incomplete penetrance, genetic heterogeneity and almost certainly multiple disease genes (Cornelia and Jorg, 2000). Epidemiological studies have found considerable variations in the incidence and prevalence of RA across time periods and geographic regions. Many factors, both genetic and environmental, such as climate, ultraviolet (UV) exposure and diet could contribute to rate of RA prevalence. Evidence from twin (Aho et al. 1986; Deighton and Walker, 1991) and family studies (MacGregor et al. 2000) suggests that both genetic and environmental factors contribute to susceptibility to rheumatoid arthritis, and disease heritability has been estimated to be 60% (Dib et al. 1996).

In this review we describe insights gained into the pathogenesis of Rheumatoid Arthritis (RA) by the techniques of modern genetics, in particular evidence from genome-wide association (GWA) studies, which provide support for the existence of a common genetic risk basis to several diseases.

The major histocompatibility complex (MHC)

The major histocompatibility complex (MHC) region on chromosome 6 contributes to the risk of almost all autoimmune diseases and its role in immunity in mice was recognized over 60 years ago. In humans, the MHC locus is also known as the HLA (human leukocyte antigen) region, reflecting the initial identification of MHC gene products on the surface of white blood cells. The classical MHC extends over around 4 megabases and comprises three clusters: class I, II, and III. Class I and II regions include genes that encode the α- and β-chains of the MHC I and II complexes, and flank the class III region, which contains an assortment of immunologically relevant genes. Despite extensive study, the mechanisms that link the MHC to disease are largely unknown, although it is supposed that variation in the MHC peptide binding cleft facilitates presentation of self-antigen to autoreactive lymphocytes.

These difficulties in understanding the MHC are not without reason; it contains some of the most polymorphic loci described in the genome and has a highly complicated genetic architecture with some regions exhibiting extended linkage disequilibrium (Horton et al. 2004).

In RA, the MHC accounts for around a third of the genetic liability (Fernando et al. 2008). Alleles at HLA-DRB1 contribute much of this risk. Additional loci contributing to the risk of RA identified by high-density genotyping include HLA-DP in patients with anticyclic citrullinated peptide antibodies (Ding et al. 2009). Again, further work is required to definitively implicate this gene rather than variants in linkage disequilibrium.

Tumor necrosis factor (TNF)

TNF-associated signalling pathway genes play a prominent role in the risk for both Systemic Lupus Erythematosus (SLE) and RA. Associations of variants in TNFAIP3, and the TRAF1-C5 locus have been identified (Ding et al. 2009; Plenge et al. 2007). TNFα-induced protein-3 (TNFAIP3) is a ubiquitin editing enzyme that acts as a negative regulator of NFκB. This can disassemble Lys63-linked polyubiquitin chains from targets such as TRAF6 and RIP1. A second region of TNFAIP3 catalyses Lys48-linked ubiquitination that targets the molecule for degradation by the proteasome (Komander and Barford, 2008). TNFAIP3 modifies key mediators in the downstream signalling of TLRs that use MyD88, TNF receptors, the IL-1 receptor family, and nucleotide-oligomerization domain protein 2 (NOD2) (Sun, 2008). Tnfaip3 knockout mice develop severe multi-organ

Corresponding Author: Dr. Peter John NUST Publishing, © (2010), ISSN: 2072-4659

inflammatory disease, and the phenotype is lethal (Lee et al. 2000). The SNP rs10499194 in TNFAIP3 carries an Odd Ratio (OR) of 1.33 for RA, and rs5029939 an OR of 2.29 for SLE (Lee et al. 2000), the latter also conferring an increased risk of haematologic or renal complications (Bates, 2009).

On chromosome 9 the region containing TRAF1 (TNF receptor associated factor 1) and C5 (complement component 5) genes is associated with significant risk for RA .TRAF1 is principally expressed in lymphocytes and inhibits NFκB signalling by TNF. This pathway is blocked in TRAF1 overexpression (Carpentier and Beyaert, 1999) whilst, conversely, Traf1-/- mice are sensitized to TNF and have exaggerated TNF-induced skin necrosis (Tsitsikov et al, 2001). The complement system has long been known to be involved in the pathogenesis of RA. In the collagen-induced arthritis model of RA, C5 deficiency prevents disease de novo and ameliorates existing symptoms and signs (Tsitsikov et al, 2001; Wang et al. 2000). Interestingly, GG homozygotes at the TRAF1-C5 SNP rs3761847 with RA have a significantly increased risk of death (hazard ratio 3.96, 95% confidence interval 1.24 to 12.6, P = 0.02) from malignancy or sepsis, potentially allowing identification of patients for appropriate screening (Panoulas et al. 2009).

Protein tyrosine phosphates, Non-receptor type 22 (PTPN22)

Outside the HLA region, the first reproducible genetic association for RA came with the implication of PTPN22 from a candidate gene approach (Begovich et al. 2004) based on linkage analysis identification of a susceptibility locus at 1p13 (Jawaheer et al. 2003). It has remained the strongest and most consistent association mapped by GWA studies in RA. A role in SLE has also been identified (Harley et al. 2008). The OR for the risk allele is around 1.75 in RA, and 1.5 in SLE. However, it should be noted that this allele (encoding the R620W mutation) is monomorphic or not disease associated in Korean or Japanese patients (Lee et al. 2009; Ikari et al. 2006). PTPN22 encodes lymphoid tyrosine phosphatase (LYP), a protein tyrosine phosphatase that inhibits T cell receptor signalling, decreasing IL-2 production. The disease associated SNP is responsible for a change from arginine to tryptophan at position 620, which inhibits binding to the SH3 domain of carboxy-terminal Src kinase. This in turn appears to enhance dephosphorylation of tyrosine residues in the Src family kinases Lck, FynT, and ZAP-70 (Cloutier and Veillette, 1999; Gjorloff-Wingren et al. 1999). The overall effect of the mutation is a reduction in T cell receptor signalling. The pathogenic effect of this is unclear, but may relate to impaired negative selection in the thymus, or lead to a reduction in regulatory T cells (Vang et al. 2008). Conversely, the R623Q variant of PTPN22, which is a loss-of-function mutation affecting the phosphatase activity of LYP, is protective against SLE (Orru et al. 2009).

Polarization towards TH1 and TH17 phenotypes: STAT4 and IL23

STAT4 encodes signal transducer and activation of transcription factor-4, responsible for signalling by IL-12, IL-23, and type 1 IFNs (Watford et al. 2004). STAT4 polarizes T cells towards TH1 and TH17 phenotypes, which has the potential to promote autoimmunity (Mathur et al. 2007). In RA the OR for the risk allele of SNP rs7574865 is 1.32 in one case-control study (Remmers et al. 2007), with a less strong disease association at rs11893432 in a meta-analysis of GWA studies (OR 1.14) (Raychaudhuri et al. 2008). There is convincing evidence that STAT4 is a risk locus for SLE in multiple racial groups (Hom et al. 2008; Namjou et al.2009), and it may be theorized that interference in type I IFN signalling may be the underlying pathogenic mechanism in this case.

Peptidyl arginine deiminase-4 (PADI4)

Peptidyl arginine deiminase-4 (PADI4) is a member of the enzyme family responsible for the post-translational citrullination of arginine residues in RA synovium, subsequently recognized by anti-cyclic citrullinated protein antibodies. In Japanese [87] and Korean patients (Kang et al. 2004), case-control association studies have identified functional haplotypes of PADI4 conferring risk of RA. However, in Caucasian populations this association is inconsistent (Gandjbakhch et al. 2009; Martinez et al. 2005)

Disease burden in Pakistan

The general prevalence of RA is estimated to be 0.5-1.0% worldwide (Caporali et al. 2009) and 0.1-0.2 % in Pakistani population (Baig, 2003). RA have been reported with different prevalence in different ethnic groups such as in European and American populations RA is more prevalent as compared to Asians (Abdel-Nasser et al. 1997; Lawrence et al. 1998; Gabriel et al. 1999; Alamanos and Drosos, 2005). The high prevalence of RA in Native American populations is up to 6.8 % while it's low in Asian countries (~0.3 %) (Silman and Hochberg, 2001; Akar et al. 2004).

In Pakistan, no work has been done on RA so far at molecular level. Although very little have been contributed to prevalence and epidemiology of rheumatic diseases. Studies performed by A. Farooqi and T. Gibson (1995; 1998) found the low prevalence of major rheumatic diseases in northern Pakistan and high prevalence of rheumatoid arthritis in affluent and poor urban communities of Pakistan.

Nadia et al (2008) indicated that the high frequency of anxiety and depression among patients with common rheumatic disorders and determine the possible relationship of different demographic and clinical variables with anxiety and depression.

Several studies indicate that depression occurs in 13-20% of patients suffering from Rheumatoid Arthritis (RA). By conservative estimates, major depression is two to three times more common in patients with RA than in the general population (Regier et al.1988). Depression increases the burden of RA to the patient and society,

increases worry about the disease and leads to more physical symptoms (DiMatteo et al. 2000; Dickens et al. 2001).

In South Pakistan, the prevalence of rheumatoid arthritis is said to be 0.9/1000 and 1.98/1000 in poor and affluent districts respectively, whereas in North Pakistan, the prevalence of major rheumatic disorders is quoted as 148/1000 (Hameed et al. 1995; Farooqi et al. 1998). Few studies have been carried out in Pakistan to look at the psychiatric morbidity in patients suffering from chronic rheumatic diseases. Information about HLA distributions and their associations with RA among people of South Asian derivation suggest marked ethnic heterogeneity (Malaviya et al 1983).

There are no data from Pakistan and whether or not such genetic influences contribute to the low prevalence of RA determined by us cannot be stated. There are two main reasons for incompleteness of the existing data on RA incidence and prevalence.

First, figures from Pakistan are scanty and need to be improved to allow understanding of intra- and inter-regional variability of RA. Second, several data suggest that both epidemiological and clinical features of RA vary over time indicating the involvement of different genes. Therefore, molecular studies are needed in the same area to identify changing patterns of the disease.

Future prospectus

Molecular characterization of genes involved in rheumatoid arthritis from in Pakistani population is required, which will describe the pathogenesis of RA at molecular level and to explore the possible relationship of the different demographic and clinical variables associated with it.

Apart from genetic factors, various viruses have been implicated in the cause and pathogenesis of rheumatoid arthritis (RA). Hepatitis C virus (HCV) infection, which has been recognized as a cause of some autoimmune diseases and which has been described as sometimes presenting with rheumatic manifestations indistinguishable from RA, might be a candidate (Maillefert et al. 2002; Hsu et al. 2003) Therefore, data needs to be generated in this regard to study the association of hepatitis C virus with RA in our population.

REFERENCES

1. Abdel-Nasser AM, Rasker JJ and Valkenburg HA. 1997. Epidemiological and clinical aspects relating to the variability of rheumatoid arthritis. Semin Arthritis Rheum. 27:123-140.
2. Aho K, Koskenvuo M, Tuominen J et al. 1986. Occurrence of rheumatoid arthritis in a nationwide series of twins. J Rheumatol,13:899-902.
3. Akar S, Birlik M, Gurler O et al. 2004. The prevalence of rheumatoid arthritis in an urban population of Izmir-Turkey. Clin Exp Rheumatol, 22:416-420.
4. Alamanos Y and Drosos AA. 2005. Epidemiology of adult rheumatoid arthritis. Autoimmun Rev, 4:130-136.
5. Baig JA. (2003). Role of Methotrexate in Rheumatoid Athritis. PhD. Thesis, Dept. of Biochemistry. University of Karachi. Pakistan.
6. Bates JS, Lessard CJ, Leon JM et al. 2009. Meta-analysis and imputation identifies a 109 kb risk haplotype spanning TNFAIP3 associated with lupus nephritis and hematologic manifestations. Genes Immun, 10:470-477.
7. Begovich AB, Carlton VE, Honigberg LA et al. 2004. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet, 75:330-337.
8. Caporali R, Pallavicini FB, Filippini, M et al. 2009. Treatment of rheumatoid arthritis with anti-TNF-alpha agents: A reappraisal, autoimmunity Reviews, 8: 274-280.
9. Carpentier I and Beyaert R. 1999. TRAF1 is a TNF inducible regulator of NF-kappaB activation. FEBS Lett, 460:246-250.
10. Cloutier JF and Veillette A. 1999. Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. J Exp Med, 189:111-121.
11. Cornelia MW and Jorg JG. 2000. Association of MHC and rheumatoid arthritis: HLA polymorphisms in phenotypic variants of rheumatoid arthritis. Arthritis Res, 2:212-216.
12. Cornelis F, Faure S, Martinez M et al. 1998. New susceptibility locus for rheumatoid arthritis suggested by a genomewide linkage study. Proc Natl Acad Sci USA, 95:10746-10750.
13. Deighton CM and Walker DJ. 1991. The familial nature of rheumatoid arthritis. Ann Rheum Dis, 50:62-65.
14. Dib C, Faure S, Fizames C et al. 1996. A comprehensive genetic map of the human genome based on 5264 microsatellites. Nature, 380:152-154.
15. Dieude P and Cornelis F. 2005. Genetic basis of rheumatoid arthritis. Joint Bone Spine, 72:520-526.
16. Ding B, Padyukov L, Lundström E et al. 2009. Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region. Arthritis Rheum, 60:30-38.
17. Ding B, Padyukov L, Lundström E et al. 2009. Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region. Arthritis Rheum, 60:30-38.
18. Fernando MM, Stevens CR, Walsh EC et al. 2008. Defining the role of the MHC in autoimmunity: a review and pooled analysis. PLoS Genet, 4:e1000024.
19. Gabriel SE, Crowson CS and O'Fallon WM. 1999. The epidemiology of rheumatoid arthritis in Rochester, Minnesota, 1955–1985. Arthritis Rheum., 42: 415-420.
20. Gandjbakhch F, Fajardy I, Ferré B et al. 2009. A functional haplotype of PADI4 gene in rheumatoid

arthritis: positive correlation in a French population. J Rheumatol, 36:881-886.
21. Gjörloff-Wingren A, Saxena M, Williams S et al. 1999. Characterization of TCR-induced receptor-proximal signaling events negatively regulated by the protein tyrosine phosphatase PEP. Eur J Immunol, 29:3845-3854.
22. Harley JB, Alarcón-Riquelme ME, Criswell LA et al. 2008. Genomewide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet, 40:204-210.
23. Hom G, Graham RR, Modrek B et al. 2008. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med, 358: 900-909.
24. Horton R, Wilming L, Lovering RC et al. 2004. Gene map of the extended human MHC. Nat Rev Genet, 5:889-899.
25. Ikari K, Momohara S, Inoue E et al. 2006. Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population. Rheumatology (Oxford), 45:1345-1348.
26. Jawaheer D, Seldin MF, Amos CI et al. 2003. Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. Arthritis Rheum, 48:906-916.
27. Kang N, Clarke AJ, Nicholson IA et al. 2004. Chard RB: Circulatory arrest for repair of postcoarctation site aneurysm. Ann Thorac Surg, 77:2029-2033.
28. Komander D and Barford D. 2008. Structure of the A20 OTU domain and mechanistic insights into deubiquitination. Biochem J, 409:77-85.
29. Lawrence RC, Helmick CG, Arnett FC et al. 1998. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum., 41:778-799.
30. Lee EG, Boone DL, Chai S et al. 2000. Failure to regulate TNF-induced NF-kappaB and cell death responses in A20-deficient mice. Science, 289:2350-2354.
31. Lee HS, Korman BD, Le JM et al. 2009. Genetic risk factors for rheumatoid arthritis differ in Caucasian and Korean populations. Arthritis Rheum, 60:364-371.
32. MacGregor AJ, Snieder H, Rigby AS et al. 2000. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum, 43:30-37.
33. Martinez A, Valdivia A, Pascual-Salcedo D et al. 2005. PADI4 polymorphisms are not associated with rheumatoid arthritis in the Spanish population. Rheumatology (Oxford), 44:1263-1266.
34. Mathur AN, Chang HC, Zisoulis DG et al. 2007. Stat3 and Stat4 direct development of IL-17-secreting Th cells. J Immunol, 178:4901-4907.
35. Namjou B, Sestak AL, Armstrong DL et al. 2009. Highdensity genotyping of STAT4 reveals multiple haplotypic associations with systemic lupus erythematosus in different racial groups. Arthritis Rheum, 60:1085-1095.
36. Orru V, Tsai SJ, Rueda B et al. 2009. A loss-of-function variant of PTPN22 is associated with reduced risk of systemic lupus erythematosus. Hum Mol Genet, 18:569-579.
37. Panoulas VF, Smith JP, Nightingale P et al. 2009. Association of the TRAF1/C5 locus with increased mortality, particularly from malignancy or sepsis, in patients with rheumatoid arthritis. Arthritis Rheum, 60:39-46.
38. Plenge RM, Seielstad M, Padyukov L et al. 2007. TRAF1-C5 as a risk locus for rheumatoid arthritis - a genomewide study. N Engl J Med, 357:1199-1209.
39. Raychaudhuri S, Remmers EF, Lee AT et al. 2008. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat Genet, 40:1216-1223.
40. Remmers EF, Plenge RM, Lee AT et al. 2007. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med, 357:977- 986.
41. Silman AJ and Hochberg MC. 2001. Epidemiology of the Rheumatic Diseases, 2nd Ed. New York: Oxford University Press, 31-71 pp.
42. Sun SC. 2008. Deubiquitylation and regulation of the immune response. Nat Rev Immunol, 8:501-511.
43. Tsitsikov EN, Laouini D, Dunn IF et al. 2001. TRAF1 is a negative regulator of TNF signaling: enhanced TNF signaling in TRAF1-deficient mice. Immunity, 15:647-657.
44. Vang T, Miletic AV, Arimura Y et al. 2008. Protein tyrosine phosphatases in autoimmunity. Annu Rev Immunol, 26:29-55.
45. Wang Y, Kristan J, Hao L et al. 2000. A role for complement in antibody-mediated inflammation: C5-deficient DBA/1 mice are resistant to collagen-induced arthritis. J Immunol, 164:4340-4347.
46. Watford WT, Hissong BD, Bream JH et al. 2004. Signaling by IL-12 and IL-23 and the immunoregulatory roles of STAT4. Immunol Rev, 202:139-156.

Molecular Modeling of Cytochrome P450 1a1 using the newly Crystallized Template Structure of CYP1A2-Human

Rehan Zafar, Fahed Parvaiz, Babar Aslam, Umar Niazi
NUST Center of Virology & Immunology
rehanzfr@gmail.com, mfahed85@gmail.com
BABARMED1@yahoo.com, saint_6671@yahoo.com

Abstract

Cytochrome family 1 enzymes play significant roles in carcinogenesis and xenobiotic detoxification. CYP1A1 is the P450 family 1 enzyme preferably expressed extrahepatically and participates extensively in monooxygenase activity which can either change the substrate to normal or carcinogenic metabolites, having the ability to initiateoncogenesis in lung and breast. Variegated structural properties evident in the prosites of available Cytochrome P450 (CYP) structures show versatility among CYP catalyzed reactions. In order to understand the CYP1A1 functions, hypothesized homology model has been constructed and characterization of the active site was performed by identifying important residues using docking studies and pharmacophore analysis. Model of CYP1A1-Human has been constructed using the available crystal structure of CYP1A2-Human. Active site and entry site of CYP1A1 was found to be more compact than CYP1A2. Difference of wild type CYP1A1 against its polymorphisms shows the role of mutations in the active site architecture, which explain that M2 and M4 mutations in CYP1A1 have no possible significant roles in the substrate binding and orientation for detoxification or carcinogenic activation. Different ligands including Alpha- naphthoflavone (ANF), Ethoxyresorufin, Theophylline, Tamoxifen, Ethanol, Phenacetin and Hesperetin were docked and reconfirm the ligand specific wet lab studies.

Key words: CYP1A1 , Homology Modelling, CYP1A2.

Introduction

Eighteen families of Cytochrome P450 and 43 subfamilies are ultimately translated from 57 genes of human with more than 58 pseudogenes (Nelson DR *et.al.,* 2004). They are the part of a multi-component electron transfer chains, called Microsomal Cytochrome P450-containing systems. P450s constitute multigenic superfamily of hemoproteins playing their biological roles as heme-thiolate monooxygenases enzymes in the synthesis and breakdown of endogenous compounds such as synthesis and metabolism of steroids and bile acids, metabolism of vitamin D and synthesis of cholesterol. They are also involved in detoxification of exogenous compounds such as hydrophobic xenobiotics (Nebert 1991) and activation of procarcinogenic to carcinogenic compounds, For example, the oxidation of benz(a)pyrene in cigarette smoke is catalysed by CYP1A1 to form BP-7,8-epoxide, which can be further oxidized by epoxide hydrolase to form 7,8-dihydrodiol (Shimada T and Oda Y 2001). Finally CYP1A1 catalyses this intermediate to form 7,8-dihydrodiol-9,10-epoxide, which is the ultimate carcinogen.(Beresford AP 1993; Guengerich FP 1995; Kawajiri K and Hayashi SI 1996). Such compounds form adducts with the DNA and proteins causing disruption of their normal physiological role (Rendic S and Di Carlo FJ 1997). [Some families of CYPs have the ability to metabolize multiple substrates, which accounts for their central role in drug to drug interactions.

In mammals, family 1 contains three well characterized P450s; CYP1A1, CYP1A2, and CYP1B1 involved extensively in the biotransformation of xenobiotics to more polar form for efficient excretion. Versatile nature of these enzymes is their capacity to oxidize multiple Polynuclear Aromatic Hydrocarbons (PAHs) (Yano JK, Hsu MH *et al.* 2005). Their induction is mediated by a ligand-activated transcription factor, the aryl hydrocarbon receptor having a basic-loop-helix PAS domain protein which binds to enhancers flanking the CYP1A1, CYP1A2 and CYP1B1 genes and stimulates their transcription.

CYP1A1 is also known as aryl hydrocarbon hydroxylase (AHH).. CYP1A1 is a P450 comprised of 512 amino acids. Its subcellular locations are endoplasmic reticulum and microsome membrane specifically expressed in lung, lymphocytes and placenta. Due to the specificity for tissues other than liver, CYP1A1 is considered to be extrahepatic. CYP1A1 mRNA levels are depressed by inflammatory cytokines (including interleukin-6 (Il-6), interleukin-1 alpha (Il-1 alpha), and tumor necrosis factor-alpha (TNF-alpha)) and growth factors (Barker CW *et.al.* 1992). A common reason for the repression of CYP1A1 could be the involvement of reactive oxygen species (ROS) (Morel Y and Barouki R 1998). Fluoroquinolones and macrolides can inhibit the expression of CYP1A1. CYP1A1 inhibitors like aryl hydrocarbon receptor (AhR) antagonist, resveratrol, in red wine can have anti-cancer nutrition activity (Casper RF *et.al.* 1999).

Several alleles have been identified of CYP1A1 gene include *1A1*1, 1A1*2A* (T>C; M1), *1A1*2B, 1A1*2C(A>G; M2), 1A1*3(T>C; M3), 1A1*4(C>A; M4) and 1A1*5 to 11* (Ingelman-Sundberg M *et.al.* 2001).*Translated products of some of which have high inducible AHH activity.*

Two polymorphisms i.e. *M1* and *M2* have been

Corresponding Author: Rehan Zafar

studied extensively in relation to cancer susceptibility and associated with high risk of smoking-induced lung cancer in Asians but not in Caucasians (Kawairi *et.al.* 1990). Recently *1A1*2B* and *M2* were found for causing increased risk of breast cancer (Moreno M *et.al.* 2008. These polymorphisms or genetic variability can be important contributors to inter individual preferences in drug biotransformation and different abilities to metabolize drugs, leading to extensive or slow metabolizers. But recent studies made an ambiguity for the mutation specific carcinogenesis activity of CYP1A1(Persson I *et.al.* 1997)

CYP1A1 have 73% amino acids sequence identity to CYP1A2, but both are susceptible to different substrates in terms of specifity and inhibition(Guengerich FP 1995; Kawajiri K and Hayashi SI 1996) CYP1A1 is involved in metabolising benz[a]pyrene and other polycyclic aromatic hydrocarbons (PAHs) to their toxic products (Guengerich FP 1995; Kawajiri K and Hayashi SI 1996; Shou M et.al. 1996), while CYP1A2 preferably oxidizes heterocyclic and aromatic amines (Hammons GJ *et.al.* 1997; Turesky RJ *et.al.* 1998).

The characterization of CYP1A1 structure will contribute to greater extent in understanding the enzyme function and modes of catalysis and may provide groundwork for the rational design of drugs and inhibitors.

CYP1A1 homology models based on several crystallographic templates, P450TERP, P450BM3, P450CAM and rabbit CYP2C5 have been reported (Iori F *et.al.* 2005; Szklarz GD and Paulsen MD 2002). Homology model dependent upon low sequence identity between CYP1A1 and the bacterial P450s or human CYPs 2C5, 2C8, 2C9 and CYP1A1 have also been reported (Lewis BC and Mackenzie PI 2007), but are unlikely to provide an accurate representation of the human CYP1A1 active-site.

This report attempts to provide the homology model and characterization of wild type human CYP1A1 and its M2 and M4 polymorphs. Different ligands including Alpha- naphthoflavone (ANF), Ethoxyresorufin, Theophylline, Tamoxifen, Ethanol, Phenacetin and Hesperetin were docked and reconfirmed for the ligand specific wet lab studies.

Experimental procedures

Homology modelling of CYP1A1

The complete sequence of the CYP1A1 was taken from UniProtKB/Swiss-Prot database (Accession # P04798) (Swiss-Prot Protein knowledgebase) (http://www.uniprot.org/). The protein sequence of CYP1A2 was 514aa long. The sequence of CYP1A1 protein was searched for reference sequence entries at NCBI (www.ncbi.nlm.nih.gov) using the algorithm PSI-BLAST from PDB database (Medha B and Aravind L 2007). Based on the distance in the evolutionary tree from CYP1A1 and the percentage sequence identity, the template structure having the PDB ID 2HI4 (CYP1A2) was chosen. Sequences of CYPs 1A1 & 1A2 were aligned using server based 3D-Coffee alignment program (O'Sullivan O 2004) and the alignment refined using RASCAL (Thompson JJD *et.al.* 2003). MODELLER release 9v7 (Eswar N *et.al.* 2006) was used for comparative modeling of the CYP1A1 structure based on the CYP1A2 template structure (PDB ID 2HI4).

Thirty two models for CYP1A1 were generated, and sixteen models of these models with lowest energies were clustered into 4 groups based on the root mean square distance (RMSD) between the corresponding residues in their structures using NMRCLUST (Kelley LA 1996). Representative 3 models from each cluster and outliers were selected for further analysis and, the free energy of the models, Ramachandran plot calculated using PROCHECK (Roman A *et.al.* 1993) and ERRAT (Colovos C and Yeates TO 1993) score were used as criterion for final model selection.

Docking of ligands:

Different ligands including Alpha-naphthoflavone (ANF), Ethoxyresorufin, Theophylline, Tamoxifen, Ethanol, Phenacetin and Hesperetin were docked in the heme neighbour. Each of these ligands was docked to generate fifty docking poses using GOLD (Abecasis GR and Cookson WO, 2000) within 15 Å of heme. The population size was 100, with a selection pressure of 1.1, 5 islands, niche size of 2, maximum operations were set to 100000 and the protein side chains were fixed. The resulting fifty docking poses for each ligand were ranked according to their corresponding binding score. The "correct" binding pose for each ligand, out of the fifty resulting poses, was chosen based on the position of the expected site of metabolism relative to the heme and the binding score. The results were then analyzed using PyMol (DeLano WL 2002) and SwissPdb Viewer (Guex N and Peitsch MC 1996).

Results and discussion

The phylogenetic tree of CYP1A1 and reference sequence entries in NCBI showed that CYP1A1 was closest to 2HI4 (CYP1A2-human), 2OJD (CYP2R1) and 2F9Q (CYP2D6); while sequence identity results showed that CYP1A1 has the highest percentage sequence identity with 2HI4 (CYP1A2) of 73%. Based on the distance in the evolutionary tree from CYP1A1 and the higher percentage sequence identity, the template structure chosen was 2HI4 (CYP1A2).

Figure 1 shows the ERRAT plot for the "best" modeled CYP1A1. It had an overall quality factor of 57.32 and had areas of high error at the N and C terminals. Additionally some areas of the active site also had a high error i.e. BC loop area (110 to 115aa), the area between the F and G helix (235 to 245aa), some residues in the I helix (315 to 320aa), Substrate recognition sites (Gotoh, 1992) (SRSs) i.e. SRS5 area (382-385aa) and SRS6 area (490 to 500aa).

The Ramachandran plots for the "best" CYP1A1 model showed good stereochemistry with 90% of residues for CYP1A1 in most favoured regions while only two residues were in disallowed regions (Ramachandran GGN *et.al.* 1963).

Human cytochrome P450 1A1 structure

The structure of modeled CYP1A1 exhibits some of the same structural properties of CYP1A2, with some structural divergence. Figure 2 shows the CYP1A1 sequence with the secondary structure elements labeled as arrows (sheets) and coils (helices)

Possible entry site

The active site of P450s is located inside the protein. The recognition of substrates thus sometimes requires dynamic changes in the protein to provide an access route as well as specific binding in the active site. Similarly an egress route through the protein provides a way for release of the product, which may be the rate-limiting step. The substrate

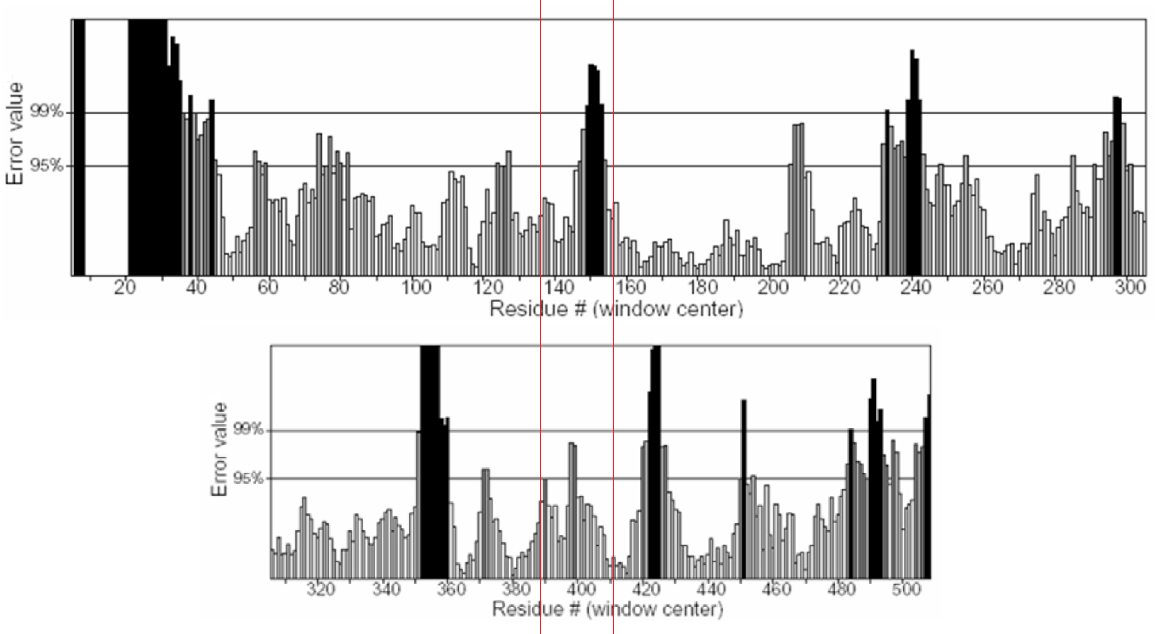

Fig. 1. ERRAT plot for CYP1A1. On the error axis (y axis), two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value. The regions with high error value (black lines) indicate areas of high energy or clashes in the structure.

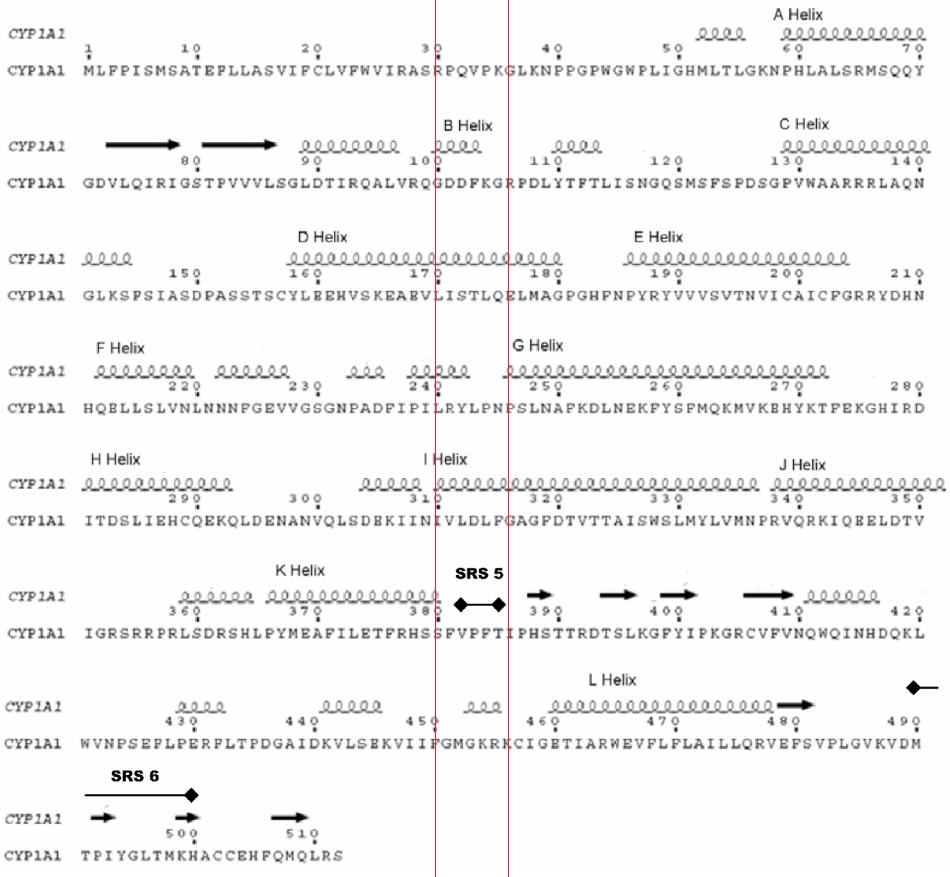

Fig. 2. Amino Acid sequence of CYP1A1 with the sheets labeled as arrows; the major helices are named and labeled as coils.

specificity of P450s can be influenced not only by the interactions made in the active site but also by accessibility to and from the active site, which highlights the importance of entrance and egress channels.

Figure 3 shows one possible entry channel into the active site. This was determined using the "Alpha site finder" utility in the program MOE (Molecular Operating Environment) (Chemical Computing group), and it corresponds to the entry channel for CYP2D6 (PDBID 2F9Q) (Rowland P *et.al.* 2006). The second potential channel appears to be between BC loop, G and I helices, similar to pw2c pathway identified by Rebecca Wade et al (Wade RC *et.al.* 2005). CYP1A2 crystal structure (PDBID 2HI4) does not show any entry channels. The entrance channel for our model appears to be lined by polar and charged residues from the F helix, I helix and E helix. The arrows show the possible entrance path. Asn219 from the F helix is replaced by Lys in CYP1A2, Asp in CYP2D6 and Ser in CYP1B1. Asn222 is replaced by His in CYP1A2, Gln in CYP2D6 and Glu in CYP1B1. Asn223 is only present in CYP1A1, while the CYPs 1B1, 1A2 and 2D6 have a Glu in this position. On the SRS6, Lys499 is conserved in CYPs 1A1, 1A2 and 1B1, while 2D6 has a SER in this position. Arg188 and Tyr187 from the E Helix also point into the entrance channel.

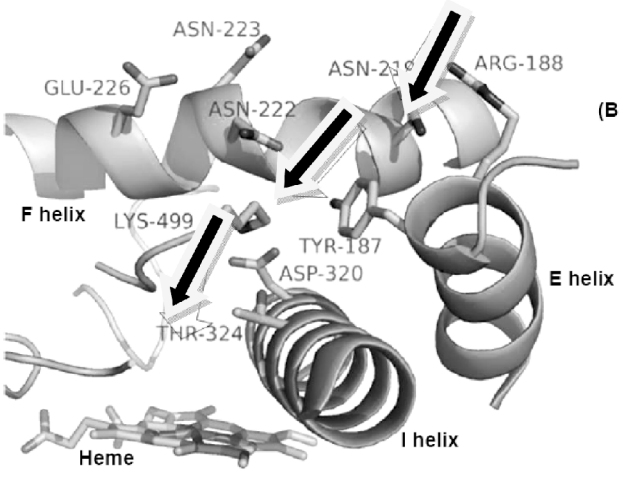

Fig. 3. Possible entry channel into the active site of CYP1A1 marked by arrows, surrounded by residues from the F helix, I helix and E helix.

Active site of CYP1A1

Active site of CYP1A1 is composed of BC loop (Figure 4), I helix (Figure 5), SRS 5 and 6 (Figure 6), F helix (Figure 7), G helix (Figure 8). The residues along the I Helix, BC Loop and SRS 5 area surround the heme, helping to place the ligand on top of the heme.

In the BC loop (Figure 4), Phe123 is conserved across a range of P450s and provides important pi stacking hydrophobic interactions placing the ligand on top of the heme. Ser122 and Ser120 can provide important hydrogen bonding interactions and Ile115 can provide hydrophobic interactions. Ser122 in CYP1A1 aligns with Thr124 in CYP1A2 (Data is not shown here). Thr124 in CYP1A2 is important in substrate binding. This was also observed in

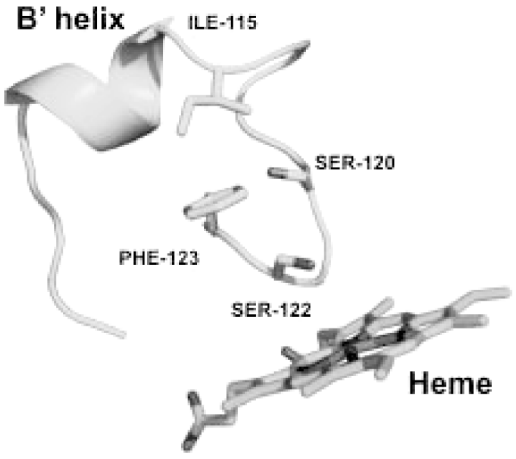

Fig. 4. Residues in the BC loop

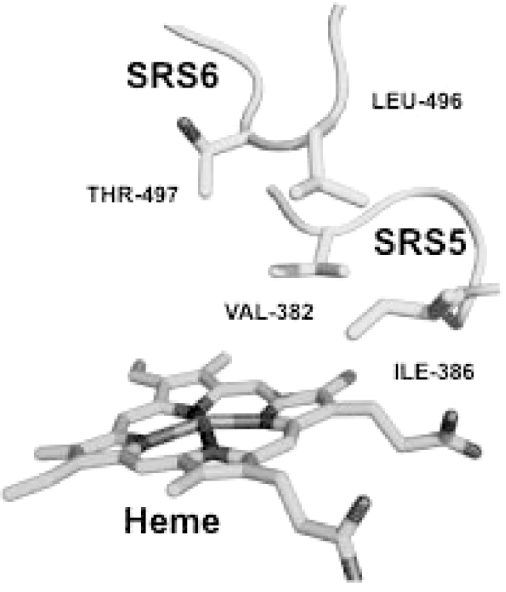

Fig. 5. Residues of I helix are oriented towards the active site in CYP1A1.

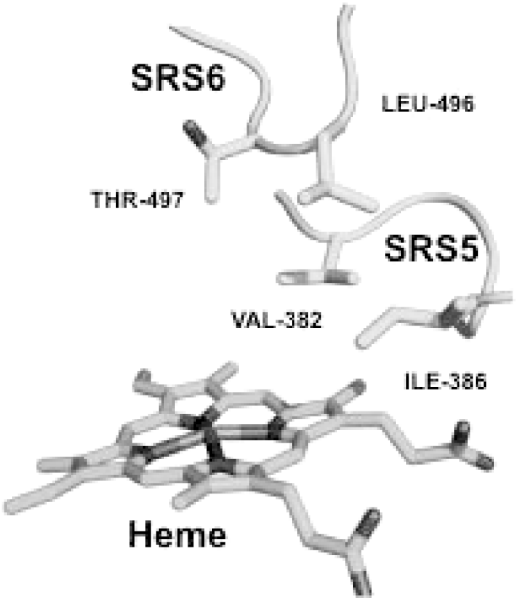

Fig. 6. Val382 and Ile386 (SRS5); Thr497 and Leu496 (SRS6) for CYP1A1 are shown.

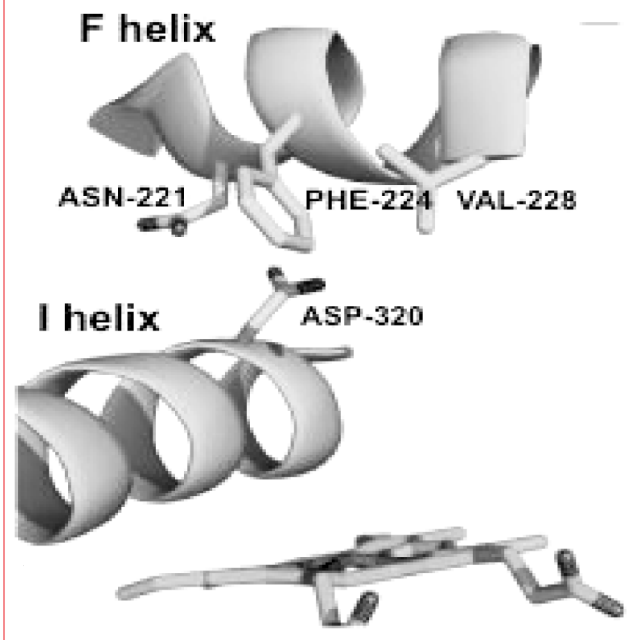

Fig. 7. Residues from F helix oriented towards the active site in CYP1A1. Asp320 (I helix) is also visible.

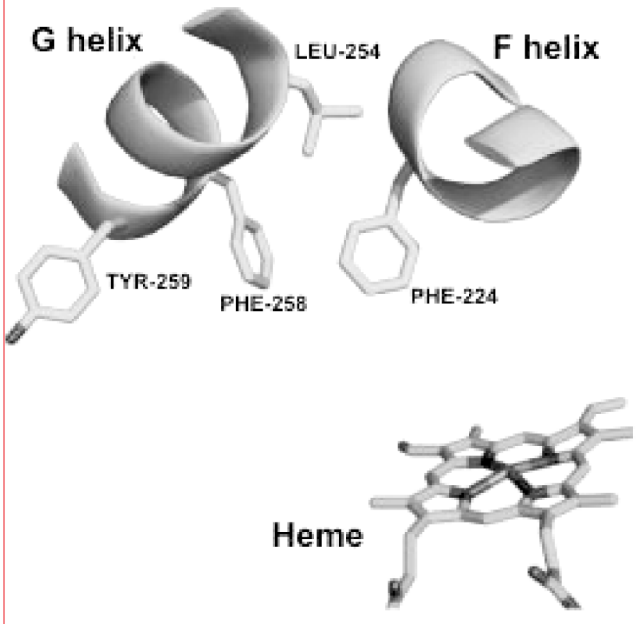

Fig. 8. Residues of G helix from CYP1A1 oriented into the active site. Phe224 from F helix is shown as well

equivalent CYP1A1 S122T mutant, which displayed significantly increased O-dealkylation activity for both 7-ethoxy and 7-methoxyresorufin over CYP1A1 wild-type (Sansen S *et. al.* 2007).

Asp313. Asp320, Ala317, Thr321 present in helix I (Figure 5) are conserved in CYPs 1A1 and 1A2.

Val382 and Ile386 (SRS5); Thr497 and Leu496 (SRS6) are shown in Figure 6. Modeling suggests that Val382 provides important hydrophobic interactions to position the ligand on top of the heme, e.g. in the case of ethoxyresorufin, the ethyl group forms hydrophobic interactions with this Val382 thus positioning the ligand on top of the heme for O-de-ethylation (EROD Activity: – Figure 12). The ethoxyresorufin-*O*-deethylase (EROD) assay is utilized for observing the induction of the CYP1A1 and is used as a biomarker for the exposure of xenobiotics that bind with to AhR (Burke M and Mayer R. 1974).

CYP1A2 has a Leu in the position of CYP1A1:Val382. Leu is a larger residue than Val, which is likely why wild-type CYP1A1 (Val-shorter residue) versus CYP1A2 shows a clear preference for 7-ethoxyresorufin versus 7-methoxyresorufin O-dealkylation. The reciprocal CYP1A1 V382L and CYP1A1 L382V mutants display interchanged specificities (Sansen SS *et.al.* 2007). Ile386 is conserved in both CYPs i.e 1A1, 1A2. Mutagenesis studies on Ile386 in CYP1A2 have shown to alter substrate specificity, e.g. I386T mutation showed a marked shift in catalytic specificity, with much reduced phenacetin O-de-ethylation but increased EROD activity (Zhou HH *et.al.* 2004). Val382, Ile386 and Leu496 appear to form an important hydrophobic network.

Phe224 present in F helix (Figure 7) provides important pi stacking interactions for the aromatic ligands, and is also conserved in CYP1A2. Asn221:1A1 (F helix) is important in determining the flexibility of CYP1A1 active site because in CYP1A2 this Asn is replaced by a Thr223 which leads to a hydrogen bond formation between Thr223 (F Helix) and Asp320 (I Helix), making the active site restricted (Sansen SS *et.al.* 2007).

In CYP1A1, Leu is present at the 254 position instead of Phe:1A2 (Figure 8). Phe258 is found conserved both in CYP1A1 and CYP1A2. Phe258 and Phe224 of F helix (Figure 8) together can form an aromatic network, sandwiching the ligand between them. Tyr259 (G helix) in CYP1A1 is present instead of a Leu (CYP1A2) in this position. It might be important, but in the current rotamer conformation it appears to be too far from the main active site cavity.

Amino acids substitutions in more than 5 polymorphisms of CYP1A1 characterize different biological responses; most of them are concerned with the activation of carcinogenic property of compounds. From previous studies, we came to know that in M2 polymorphs the Ile462 is mutated to the Val-462. The major oxidative routes of estrone and estradiol are 2- and 4-hydroxylation by cytochrome P450 2B1, 1A and 3A. The presence of polymorphs can ultimately generate 2-OH derivative of estradiol in excess which can ultimately fatal in breast cancers. The structure determination of Polymorphisms 1A1*2A (M1), 1A1*2B and 1A1*2C (M2) can thus be defended for their role in oncogenesis.

In M2, the difference of Ile-462-Val was thought as a major fact for the oncogenic biotransformation of metabolites. But Persson and collegues (Persson I *et.al.* 1997) observed that the Val-462 variant was not functionally important in the PAH induced cancerogenesis. Furthermore, Smart and Daly (Smart J and Daly AK 2000) reported that polymorphisms in the Ah receptor gene contributed more to CYP1A1 levels than did the polymorphism in the CYP1A1 protein. From our modeled CYP1A1 structure, it can be supposed that the mutation in Ile462Val would neither induce any conformational change

in the active site of CYP1A1 nor participate in the orientation of different substrates for catalysis.

Actually, the 462 position was on the downward side of chair like coordinates for active site (Figure 9). The heme group is present in y axis whereas the 462 position is with –y value in coordinates. (Figure 9). As the role of the CYP1A1, DNA sequence variants in cigarette smoking related cancer development is not yet resolved (Watanabe M 1998; Houlston RS 2000) but the model of CYP1A1 showed that the M2 polymorph has nothing to do with the active site disruption which can ultimately result in the oncogenesis.

Similarly the mutation Thr461Asn in M4 polymorph of CYP1A1 has shown that the effect of mutations should not affect the substrate binding in the active site and confirmed by the previous wet lab studies of these mutations and their substrate binding specificity or metabolisation.

Docking analysis

The single preferred orientation, observed for Alpha-naphthoflavone (ANF) binding with CYP1A2 is shown in Figure 10. ANF is metabolised by CYP1A1 to form ANF-5, 6-diol and ANF-5,6-oxide, and is docking in a pose to be metabolised at position 5 and 6 (Figure 11). The possible reasons for this difference between ANF binding in CYP1A1 and CYP1A2 could be explained by looking at s 10 and 11. CYP1A2 has a triad of Phe, while CYP1A1 is missing the Phe256, instead being replaced here by Leu254. Leu254 (CYP1A1) is about 5.6 Å away from the ANF, which makes it a very weak hydrophobic interaction; while Phe256 (CYP1A2) is about 4.2 Å away, thus having stronger hydrophobic interactions.

The second difference, which appears to be important is CYP1A1 having a Val382 on SRS5, which is smaller and creates more space for the triple ring of ANF to fit there. On the other hand CYP1A2 has Leu382 in this position, which is a longer residue. The triple ring of ANF will clash with Leu382 if it binds that way. These two differences seem to be the most obvious for ANF binding one way in CYP1A1 and the other way in CYP1A2.

The docking pose for ethoxyresorufin (Figure 12) shows a good position for O-de-ethylation to occur. Phe123 provides pi stacking interactions; Ser122 makes a hydrogen bond with the nitrogen while the ethyl group has hydrophobic interactions Val382, Ile386 and Leu495.

Theophylline (Rendic SS 2002) can be metabolised by CYP1A1 at positions 1, 3 or 8. Figure 13 shows the docking pose for theophylline in CYP1A1 model. It is in a good orientation for 1-de-methylation. Asp320 makes a hydrogen bond with the Nitrogen at position 7, while Ser122 donates a hydrogen bond to the Oxygen.

Tamoxifen is metabolised by CYP1A2 and CYP1A1 into N-desmethylTAM (Crewe HK *et.al.* 2002). Figure 14 shows the docking pose of tamoxifen into the CYP1A1 model. Phe224 and Phe258 sandwich the ring, providing strong hydrophobic interactions along with Leu254. Phe123 and Ile115 provide hydrophobic interactions along the BC loop. Close to the heme Ile386, Val382 and Leu496 form a triad of hydrophobic residues.

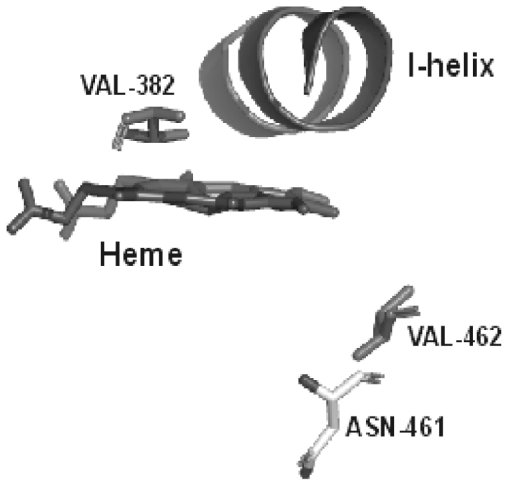

Fig. 9. Position of Asn-461 & Val-462 in M2 Polymorph of CYP1A1.

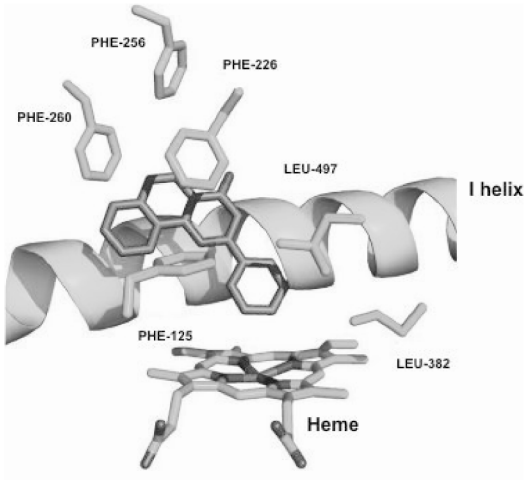

Fig. 10. ANF binding in CYP1A2 crystal structure 2HI4.

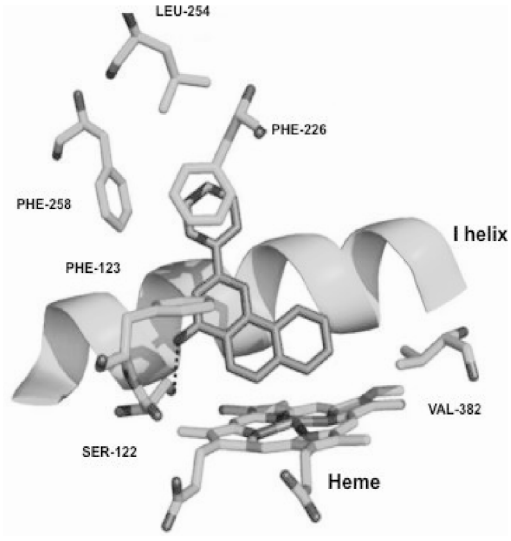

Fig. 11. ANF docking pose in the CYP1A1 model. Ser122 makes a hydrogen bond, (dashed line). Val382 (being shorter than Leu from 1A2) has more space for the triple ring of ANF to fit on top of the heme, thus placing positions 5 and 6 of ANF on top of the heme.

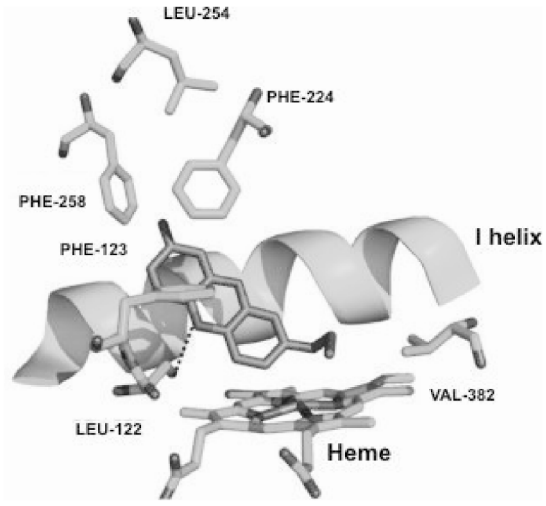

Fig. 12. Ethoxyresorufin docking into the CYP1A1 model. A possible hydrogen bond between Ser122 and Nitrogen is shown.

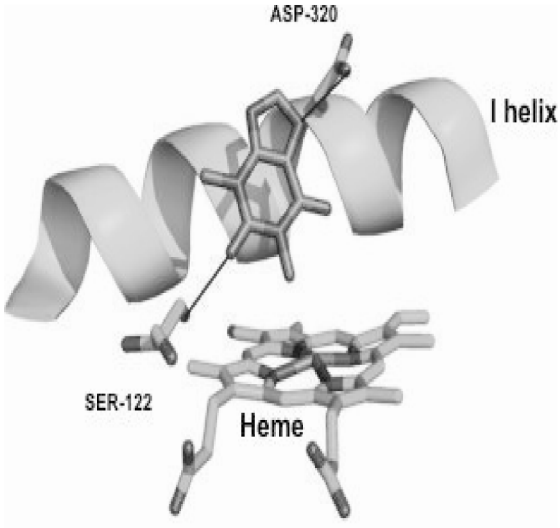

Fig. 13. Theophylline docking into CYP1A1 model in a good orientation for 1-de-methylation. Ser122 (BC loop) and Asp320 3(I helix) are involved in hydrogen bonding with Theophylline.

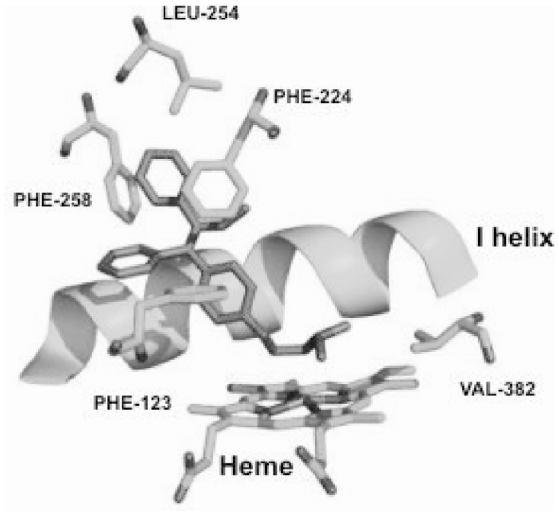

Fig. 14. Docking pose of tamoxifen into CYP1A1 model. Tamoxifen is in a good position for N-de-methylation.

Figures 15, 16 and 17 show the docking poses for some other ligands i.e. Ethanol, Phenacetin and Hesperetin (Rendic SS 2002). All three dock in good positions for metabolism to occur.

Extensive utilization of homology models to predict key amino acids that may have a role in interaction with the substrate can give an efficient throughput screening for the lead products. For Example, *In silico* docking Studies with CYP3A4 (Szklarz and Halpert, 1997; He YA *et.al.* 1997) showed important key residues which were then latterly confirmed by site-directed mutagenesis (He YA *et.al.* 1997; Domanski TL *et.al.* 1998). Our studies have identified some residues in the active site which are believed to be involved in substrate recognition and binding, including: Ile115, Ser122, Phe123, Phe224, Phe258, Tyr259, Asp313, Ala317, Asp320, Thr321, Val382, Ile386, Leu496 and Thr497. These residues provide an avenue to study CYP1A1 active site and its interaction with lead compounds.

Conclusion

A comparative model of CYP1A1 based on the structure of P450 1A2 built having 73% of sequence homology. Since, a profound comparative model always require optimum homologue template for defining its restraints, we presented a comparative model of CYP1A1 with the highest confidence in the accuracy of restraints in comparison to the earlier models which were modeled using the maximum limit of 36% homology. Such studies are commonly performed *in vitro* and *in silico* in the process of drug development to estimate the risk of drug interactions. Our CYP1A1 homology models give good insight into the CYP1A1 active site and its interaction with various substrates. Our studies have identified some residues in the active site which are believed to be involved in substrate recognition and binding, including: Ile115, Ser122, Phe123, Phe224, Phe258, Tyr259, Asp313, Ala317, Asp320, Thr321, Val382, Ile386, Leu496 and Thr497. These residues are identical to the residues identified in other studies (Szklarz GD and Paulsen MD 2002; Iori F *et.al.* 2005; Lewis BC *et.al.* 2007). The models also predict sites of metabolism for some of the ligands and binding positions of some ligands like Ethoxyresorufin which are similar to those predicted in other studies (Lewis BC *et.al.* 2007). Our model suggests the role of mutations to be ineffective for metabolisation of substrates and reconfirm the observations of Persson and collegues (Persson I *et.al.* 1997).The search for CYP1A1 inhibitors together with *in silico* models developed herein may provide useful information for the development of chemoprotectors and wet lab conformational studies.

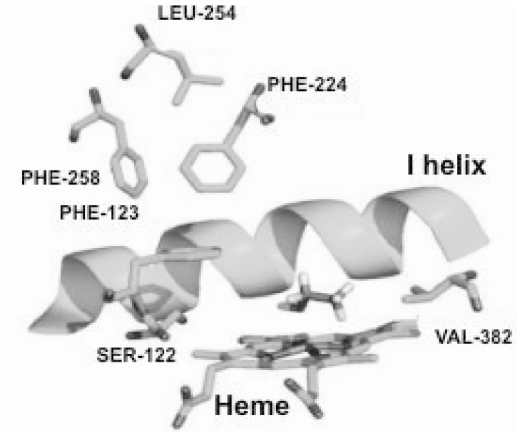

Fig. 15. Ethanol docking pose into CYP1A1 model. It is in a good position for oxidition to Acetaldehyde. The Hydrogens on Ethanol are shown for clarity.

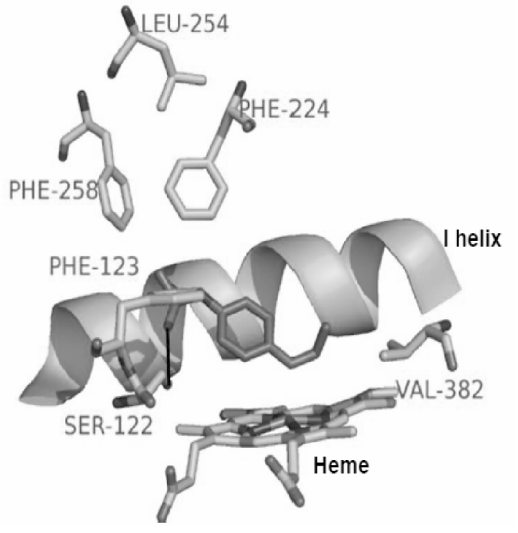

Fig. 16. Phenacetin docking pose in the CYP1A1 model. The oxygen is positioned on top of the heme for O de-ethylation to form Paracetamol. Ser122 forms a hydrogen bond and Val382 has hydrophobic interactions with the ethyl group to position the oxygen on top of the heme.

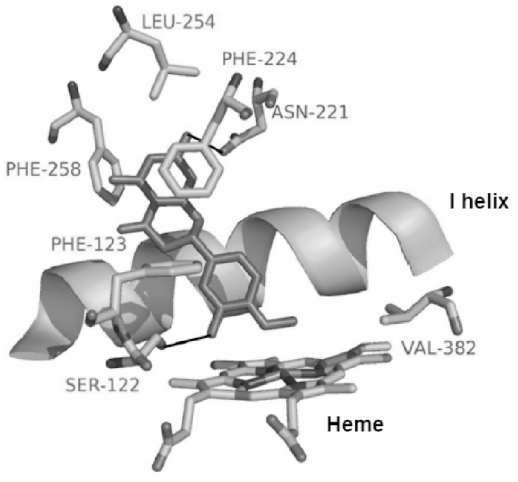

Fig. 17. Hesperetin docking pose in the CYP1A1 model. Val382 provides hydrophobic interactions with the methyl group, Ser122 and Asn221 make hydrogen bonds while Phe224 and Phe258 provide pi stacking interactions. These interactions place Hesperetin on top of the heme for O demethylation to occour.

REFERENCES

1. Abecasis GR and Cookson WO. 2000. GOLD--graphical overview of linkage disequilibrium. Bioinformatics, 16(2):182-183.
2. Barker CW, Pagan JB and Pasco DS. 1992. Interleukin-1β suppresses the induction of P450 1A1 and P450 1A2 mRNAs in isolated hepatocytes. Journal of Biological Chemistry, 267: 8050–8055
3. Beresford AP. 1993. CYP1A1: friend or foe? Drug Metabolism Reviews, 25 (4): 503–517.
4. Burke M and Mayer R. 1974. Ethoxyresorufin: direct fluometric assay of a microsomal O-dealkylation which is preferentially inducible by 3-methylchloranthrene. Drug Metabolism and Disposition, 2:583–588.
5. Casper RF, Quesne M, Rogers IM, Shirota T, Jolivet A, Milgrom E and Savouret JF .1999. Resveratrol has antagonist activity on the aryl hydrocarbon receptor: implications for prevention of dioxin toxicity. Molecular Pharmacology, 56:784-790.
6. Colovos C and Yeates TO.1993. Verification of protein structures — Patterns of nonbonded atomic interactions. Prot. Sci. 2, 1511–1519.
7. Crewe HK, Notley LM, Wunsch RM, Lennard MS, and Gillam EMJ. 2002. Metabolism of Tamoxifen by Recombinant Human Cytochrome P450 Enzymes: Formation of the 4-Hydroxy, 4'-Hydroxy and N-Desmethyl Metabolites and Isomerization of trans-4-Hydroxytamoxifen.Drug Metabolism and Disposition, 30(8): 869 - 874.
8. DeLano WL.2002. The PyMOL Molecular Graphics System. DeLano Scientific San Carlos, CA, USA. http://www.pymol.org. Accessed on 12-01-2010.
9. Domanski TL, Liu J, Harlow GR and Halpert JR. 1998. Analysis of four residues within substrate recognition site 4 of human cytochrome P450 3A4: role in steroid hydroxylase activity and α-naphthoflavone stimulation. Achieve of Biochemistry and Biophysics, 350:223-232.
10. Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A. 2006. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30: 200.
11. Gotoh .1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences, Journal of Biological Chemistry, 267: 83–90.
12. Guengerich FP. 1995. Human cytochrome P450 enzymes, in Cytochrome P450: Structure, Mechanism and Biochemistry. Plenum Press, New York, 473–535.
13. Guex N and Peitsch MC.1996. Swiss-PdbViewer: A Fast and Easy-to-use PDB Viewer for Macintosh and PC. Protein Data Bank Quaterly Newsletter, 77: 7.
14. Hammons GJ, Milton D, Stepps K, Guengerich FP, Tukey RH, Kadlubar FF.1997. Metabolism of carcinogenic heterocyclic and aromatic amines by recombinant human cytochrome P450 enzymes. Carcinogenesis, 18: 851–854.

15. He YA, He YQ, Szklarz GD and Halpert JR. 1997. Identification of three key residues in substrate recognition site 5 of human cytochrome P450 3A4 by cassette and site-directed mutagenesis. Biochemistry, 36:8831-8839.
16. Hildebrand DC, Palleroni NJ, Hendson M, Toth J and Johnson JL. 1994. Pseudomonas avescens isolated from walnut blight cankers. Systematic Bacteriology, 44: 410–415.
17. Houlston RS. 2000. CYP1A1 polymorphisms and lung cancer risk: a meta-analysis. Pharmacogenetics, 10:105–114.
18. Iori F, da Fonseca R, João Ramos M and Menziani MC. 2005. Theoretical quantitative structure-activity relationships of flavone ligands interacting with cytochrome P450 1A1 and 1A2 isozymes. Bioorganic and Medicinal Chemistry, 13(14): 4366-4374.
19. Ingelman-Sundberg M, Oscarson M, Daly AK, Garte S, Nebert DW. 2001. Human cytochrome P-450 (CYP) genes: a web page for the nomenclature of alelles. Cancer Epidemiology, Biomarkers and Prevention, 10: 1307-1308.
20. Kawajiri K and Hayashi SI. 1996. The CYP1 family: Cytochromes P450-Metabolic and Toxicological Aspects. CRC Press, Boca Raton, Florida, 77–97.
21. Kawajiri K,Nakachi K,Imai K,Yoshii A,Shinoda N,Watanabe J. 1990. Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome P450IA1 gene. FEBS Lett, 263: 131–133.
22. Kelley LA, Gardner SP and Sutcliffe MJ. 1996. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Engineering, 9: 1063-1065.
23. Lewis BC, Mackenzie PI, and Miners JO. 2007. Comparative homology modeling of human cytochrome P4501A1 (CYP1A1) and confirmation of residues involved in 7-ethoxyresorufin O-deethylation by site-directed mutagenesis and enzyme kinetic analysis. Archives of Biochemistry and Biophysics, 468(1):58-69.
24. Medha B and Aravind L.2007. PSI-BLAST Tutorial. Comparative Genomics. Methods in Molecular Biology, 395: 177-186.
25. Moreno M, Sarti E, Herrera N, Ramirez J, Robles V and Tapia R. 2008. Steroid metabolism gene CYP17, CYP1A1, 2B, CYP1A1, 2C and risk of breast cancer in Mexican women. Journal of Clinical Oncology, ASCO Annual Meeting Proceedings, 26(15S): 1524
26. Morel Y and Barouki R.1998. Down-regulation of cytochrome P450 1A1 gene promoter by oxidative stress: Critical contribution of nuclear factor 1. Journal of Biological Chemistry, 273:26969–26976.
27. Molecular Operating Environment - MOE. 2008, Chemical Computing Group Inc.
28. Nebert DW. 1991. Role of genetics and drug metabolism in human cancer risk. Mutation Research, 247(2):267-281.
29. Nelson DR, Zeldin DC, Hoffman SM, Maltais LJ, Wain HM and Nebert DW. 2004. Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. Pharmacogenetics, 14(1):1-18.
30. O'Sullivan O. 2004. Combining Protein Sequences and Structures within Multiple Sequence Alignments. Journal of Molecular Biology, 340: 385-395.
31. Persson I, Johansson I, Ingelman SM. 1997. In vitro kinetics of two human CYP1A1 variant enzymes suggested to be associated with interindividual differences in cancer susceptibility. Biochemical and Biophysical Research Communications, 231: 227–30.
32. Ramachandran GGN, Ramakrishnan CC, and Sasisekharan VV.1963. Stereochemistry of polypeptide chain configurations. Journal of Molecular Biology, 7: 95-99.
33. Rendic S and Di Carlo FJ. 1997. Human cytochrome P450 enzymes: A status report summarizing their reactions, substrates, inducers, and Inhibitors. Drug Metabolism Reviews, 29: 413-580.
34. Roman A. Laskowski, Malcolm W. MacArthur, David S. Moss and Janet M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26, 283-291.
35. Rowland P, Blaney FE, Smyth MG, Jones JJ, Leydon VR, Oxbrow AK, Lewis CJ, Tennant MG,
36. Modi S, Eggleston DS, Chenery RJ and Bridges AM .2006. Crystal Structure of Human Cytochrome P450 2D6. Journal of Biological chemistry, 281:7614-7622.
37. Rendic S. 2002, Summary of information on human CYP enzymes: human P450 metabolism data. Drug Metabolism Review, 34: 83-448.
38. Sansen S, Yano JK, Reynald RL, Schoch GA, Griffin KJ, Stout CD and Johnson EF. 2007. Adaptations for the Oxidation of Polycyclic Aromatic Hydrocarbons Exhibited by the Structure of Human P450 1A2. Journal of Biological chemistry, 282: 14348-14355.
39. Schoch GA, Yano JK, Wester MR, Griffin KJ, Stout CD and Johnson EF. 2004. Biological Chemistry, 279: 9497–9503.
40. Scott EE, Wester MR, White M A, Chin CC and Stout CD. 2003. Proceedings of the National Academy of Sciences, USA, 100: 13196–13201.
41. Shimada, T, Oda Y, Gillam EM, Guengerich FP and Inoue K. 2001. Metabolic activation of polycyclic aromatic hydrocarbons and other procarcinogens by cytochromes P450 1A1 and P450 1B1 allelic variants and other human cytochromes P450 in Salmonella typhimurium NM2009. Drug Metabolism and Disposition, 29: 1176–1182.
42. Shou M, Gonzalez FJ and Gelboin HV. 1996. Stereoselective epoxidation and hydration at the K-region of polycyclic aromatic hydrocarbons by cDNA-expressed cytochromes P450 1A1, 1A2, and epoxide hydrolase. Biochemistry, 35:15807–15813.
43. Smart J, Daly AK. 2000. Variation in induced CYP1A1 levels: relationship to CYP1A1, Ah receptor and GSTM1 polymorphisms. Pharmacogenetics, 10: 11–24.

44. Swiss-Prot Protein knowledgebase, Swiss Institute of Bioinformatics, (http://www.expasy.org/sprot) accessed on 24-12-2009.
45. Szklarz GD and Halpert JR. 1997. Molecular modeling of cytochrome P450 3A4. Journal of Computer Aided Molecular Design, 11: 265-272.
46. Szklarz GD and Paulsen MD. 2002. Molecular modeling of cytochrome P450 1A1: enzymesubstrate interactions and substrate binding affinities. Biomolecular structural Dynamics, 20:155–162.
47. Szklarz GD, Halpert JR.1997b. Molecular modeling of cytochrome P450 3A4. Journal of Computer-Aided Molecular Design, 11:265–272.
48. Thompson JJD, Thierry JJC, and Poch OO. 2003. RASCAL: Rapid scanning and correction of multiple sequence alignments. Bioinformatics, 19: 1155.
49. Turesky RJ, Constable A, Richoz J, Varga N, Markovic J, Martin MV, Guengerich FP.1998. Activation of heterocyclic aromatic amines by rat and human liver microsomes and by purified rat and human cytochrome P450 1A2. Chemical Research in Toxicology, 11: 925–936.
50. Wade RC, Motiejunas D, Schleinkofer K, Sudarko, Winn PJ, Banerjee A, Kaniakin A and Jung C. 2005. Multiple molecular recognition mechanisms. Cytochrome P450—a case study. International Journal of Biochemistry, Biophysics and Molecular Biology, 1754:239–244.
51. Watanabe M. 1998. Polymorphic CYP genes and disease predisposition. Toxicology Letters, 102-103: 167–71.
52. Wester MR, Johnson EF, Marques-Soares C, Dijols S, Dansette PM, Mansuy D, Stout CD. 2003b. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 Å resolution: evidence for an induced fit model of substrate binding. Biochemistry, 42:9335–9345.
53. Yannick M and Robert B. 2000. The repression of nuclear factor I/CCAAT transcription factor (NFI/CTF) transactivating domain by oxidative stress is mediated by a critical cysteine (Cys-427). Biochemistry, 348: 235–240.
54. Yano JK, Hsu MH, Griffin KJ, Stout CD, and Johnson EF. 2005. Structures of human microsomal cytochrome P450 2A6 complexed With coumarin and methoxsalen. Nature Structural And Molecular Biology, 12: 822–823.
55. Zhou H, Josephy PD, Kim D and Guengerich FP. 2004. Functional characterization of four allelic variants of human cytochrome P450 1A2. Archives of Biochemistry and Biophysics, 422: 23–30.

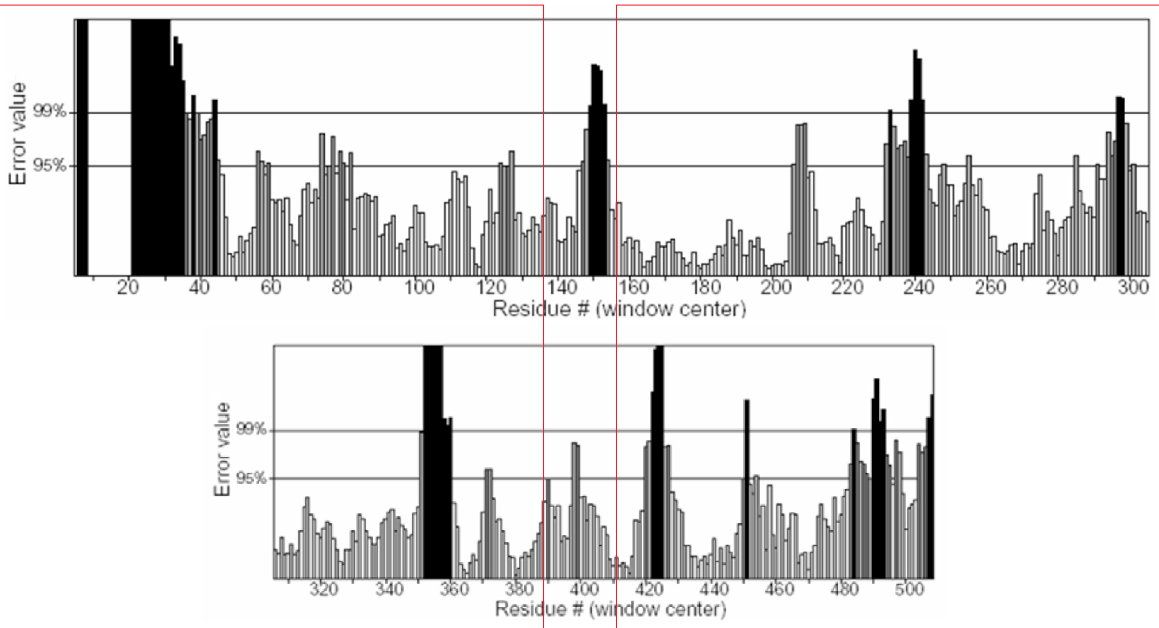

Fig. 1. ERRAT plot for CYP1A1. On the error axis (y axis), two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value. The regions with high error value (black lines) indicate areas of high energy or clashes in the structure.

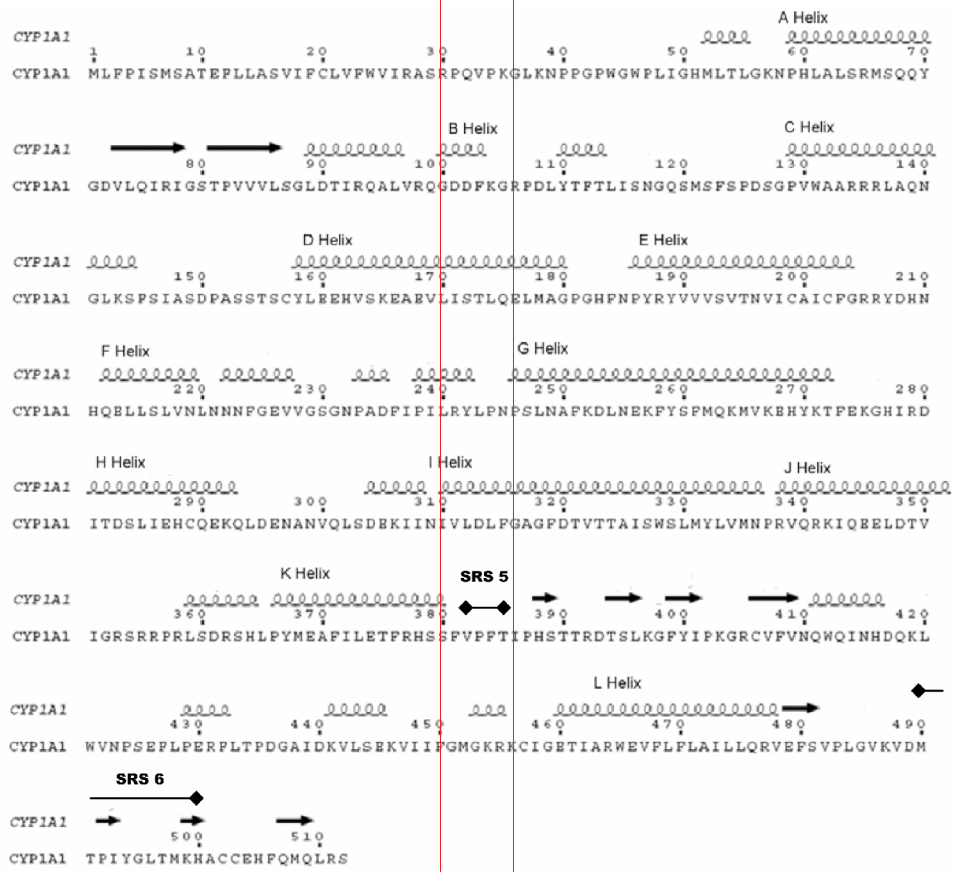

Fig. 2. Amino Acid sequence of CYP1A1 with the sheets labeled as arrows; the major helices are named and labeled as coils.

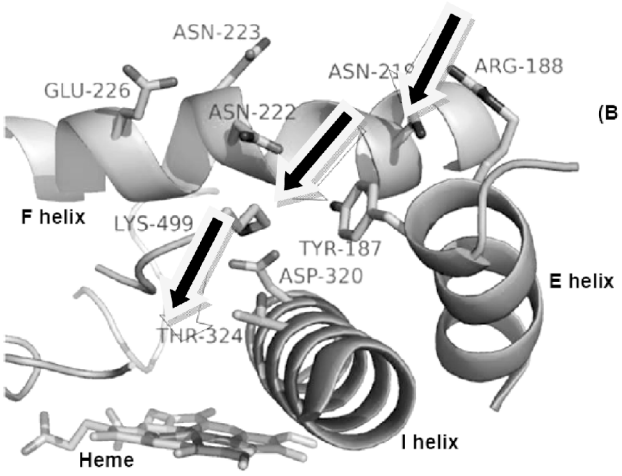

Fig. 3. Possible entry channel into the active site of CYP1A1 marked by arrows, surrounded by residues from the F helix, I helix and E helix.

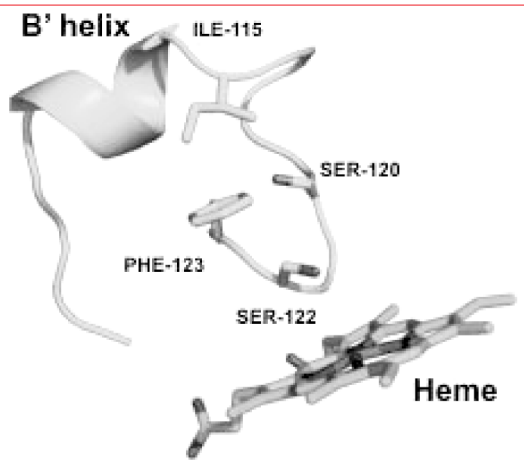

Fig. 4. Residues in the BC loop

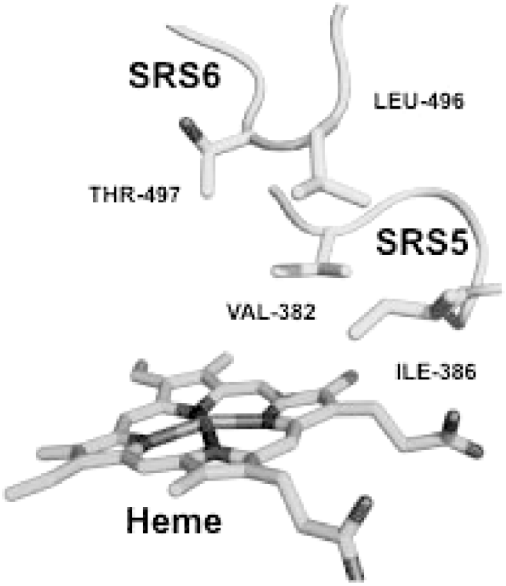

Fig. 5. Residues of I helix are oriented towards the active site in CYP1A1.

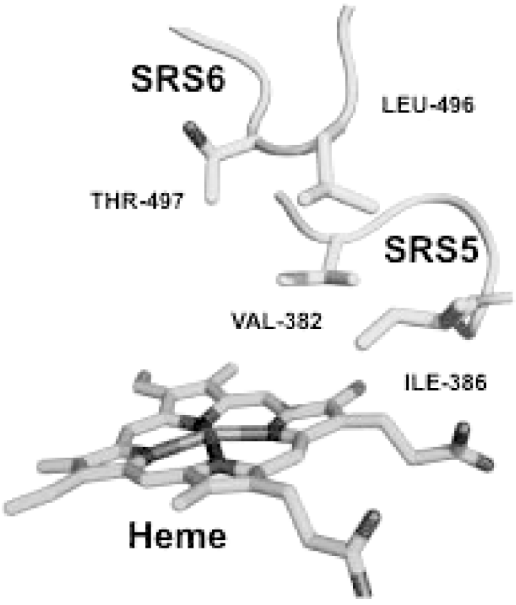

Fig. 6. Val382 and Ile386 (SRS5); Thr497 and Leu496 (SRS6) for CYP1A1 are shown.

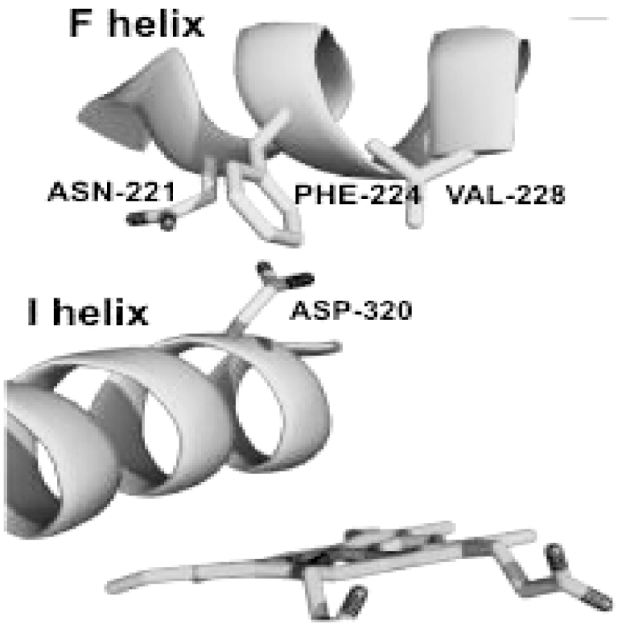

Fig. 7. Residues from F helix oriented towards the active site in CYP1A1. Asp320 (I helix) is also visible.

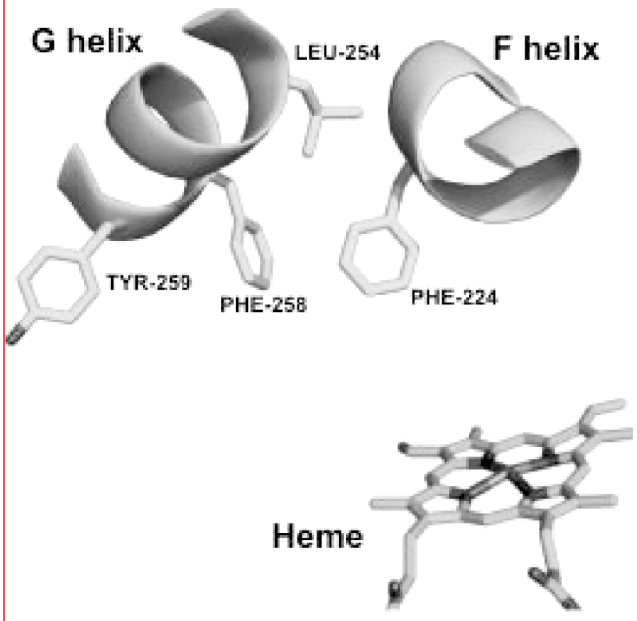

Fig. 8. Residues of G helix from CYP1A1 oriented into the active site. Phe224 from F helix is shown as well

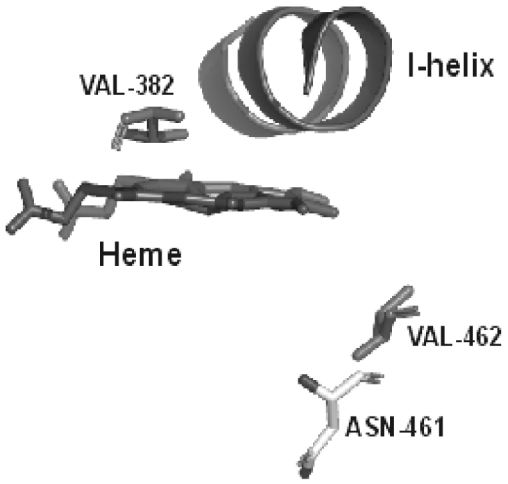

Fig. 9. Position of Asn-461 & Val-462 in M2 Polymorph of CYP1A1.

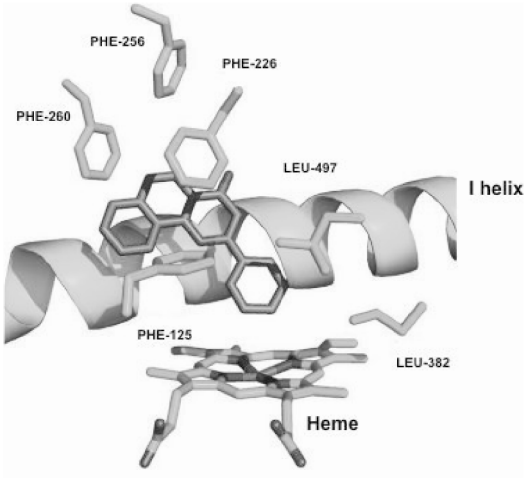

Fig. 10. ANF binding in CYP1A2 crystal structure 2HI4.

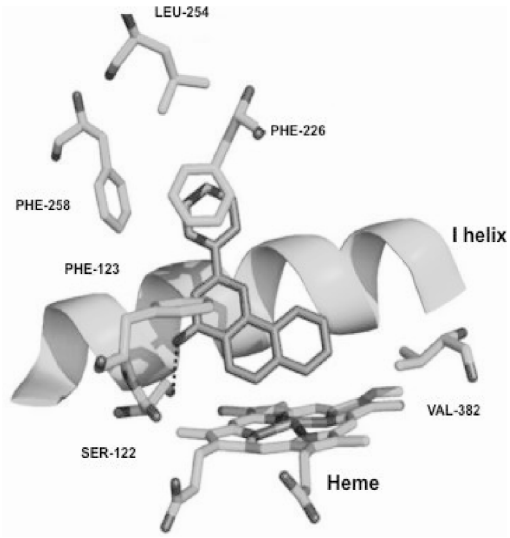

Fig. 11. ANF docking pose in the CYP1A1 model. Ser122 makes a hydrogen bond, (dashed line). Val382 (being shorter than Leu from 1A2) has more space for the triple ring of ANF to fit on top of the heme, thus placing positions 5 and 6 of ANF on top of the heme.

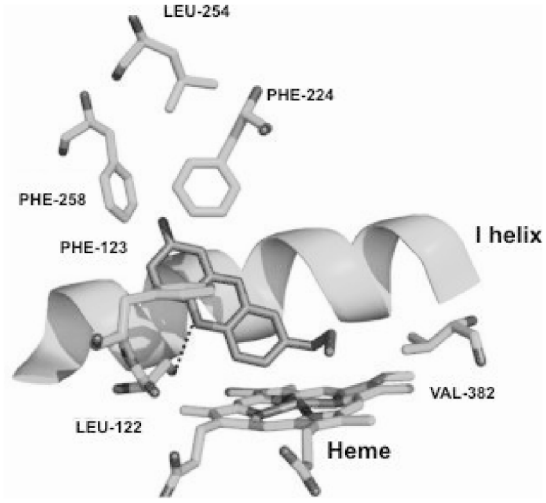

Fig. 12. Ethoxyresorufin docking into the CYP1A1 model. A possible hydrogen bond between Ser122 and Nitrogen is shown.

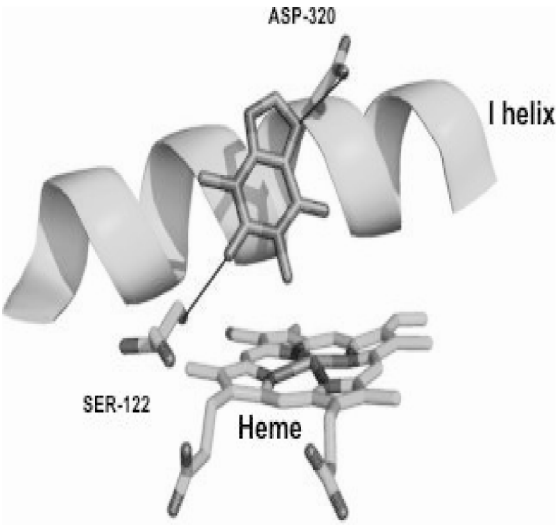

Fig. 13. Theophylline docking into CYP1A1 model in a good orientation for 1-de-methylation. Ser122 (BC loop) and Asp320 3(I helix) are involved in hydrogen bonding with Theophylline.

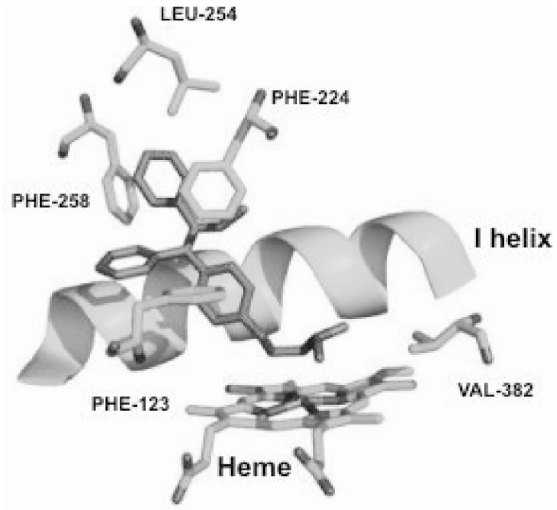

Fig. 14. Docking pose of tamoxifen into CYP1A1 model. Tamoxifen is in a good position for N-de-methylation.

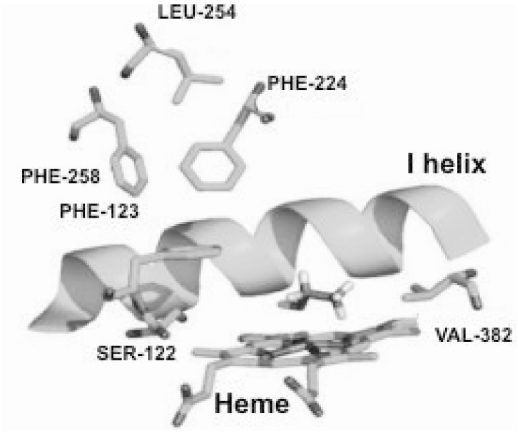

Fig. 15. Ethanol docking pose into CYP1A1 model. It is in a good position for oxidition to Acetaldehyde. The Hydrogens on Ethanol are shown for clarity.

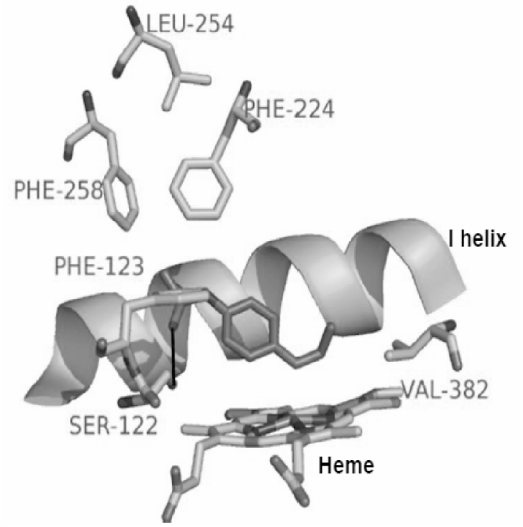

Fig. 16. Phenacetin docking pose in the CYP1A1 model. The oxygen is positioned on top of the heme for O de-ethylation to form Paracetamol. Ser122 forms a hydrogen bond and Val382 has hydrophobic interactions with the ethyl group to position the oxygen on top of the heme.

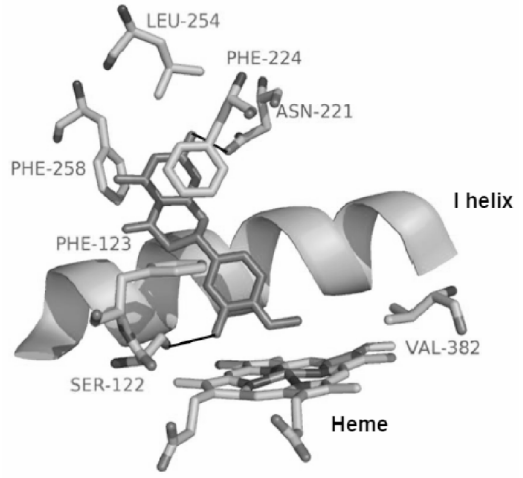

Fig. 17. Hesperetin docking pose in the CYP1A1 model. Val382 provides hydrophobic interactions with the methyl group, Ser122 and Asn221 make hydrogen bonds while Phe224 and Phe258 provide pi stacking interactions. These interactions place Hesperetin on top of the heme for O demethylation to occour.