

UNCONVENTIONAL PROTEIN TARGETS OF MULTI DRUG RESISTANT *ACINETOBACTER*

Ph.D. THESIS

by

JYOTI SINGH TOMAR



DEPARTMENT OF CHEMISTRY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247 667 (INDIA)
NOVEMBER, 2015

**UNCONVENTIONAL PROTEIN TARGETS OF
MULTI DRUG RESISTANT *ACINETOBACTER***

A THESIS

*Submitted in partial fulfilment of the
requirements for the award of the degree*

of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

JYOTI SINGH TOMAR



**DEPARTMENT OF CHEMISTRY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247 667 (INDIA)
NOVEMBER, 2015**

**©INDIAN INSTITUTE OF TECHNOLOGY ROORKEE, ROORKEE-2015
ALL RIGHTS RESERVED**

DEDICATED TO

**My parents Pratap Singh Tomar and Sudha Singh
&
My (late) Grandfather Shivrath Singh Tomar**



INDIAN INSTITUTE OF TECHNOLOGY ROORKEE ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled '**UNCONVENTIONAL PROTEIN TARGETS OF MULTI DRUG RESISTANT ACINETOBACTER**' in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy and submitted in the Department of Chemistry of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from December 2009 to November 2015 under the supervision of Dr. Rama Krishna Peddinti, Associate Professor, Department of Chemistry, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

(Jyoti Singh Tomar)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(R. K. Peddinti)
Supervisor

Date: November 16, 2015

The Ph.D. Viva-Voice Examination of **Ms. Jyoti Singh Tomar**, Research Scholar, has been held on March 07, 2016.

Signature of Supervisor

Signature of External Examiner

1.1. Introduction

Bacteria can be grouped into two categories as beneficial and harmful. The bacteria which co-exist with humans without causing any diseases are known as beneficial. Sometimes, these turn out to be dangerous when they infect people having a compromised immune system, *viz.* hospitalized patient or person suffering from cancer, HIV or diabetes. Harmful bacteria are the major cause of illness and leading to death in severe cases. Bacterial infections borne diseases are the second leading cause of death worldwide. Primary classification of bacteria is done on the basis of a laboratory test known as Gram-staining. The Gram-stain differentiates bacteria into two types, *i.e.*, Gram-positive and Gram-negative. Gram-negative bacteria would not be able to retain the stain and would appear pink under the microscope; whereas, Gram-positive bacteria would appear purple in colour due to the retention of stain by their cell wall. This happens because of the difference in the composition of their cell walls. The cell wall forms a tight barrier against the outside environment providing immunity to both these bacterial types from different medications as well as host immune responses. Secondary metabolites have been observed to neutralize this defense barrier of bacteria. First effective metabolite discovered against bacterial cell wall was penicillin, and with this discovery it declared the beginning of golden age of antibiotics. Major diseases caused by Gram-negative and Gram-positive bacteria are listed in Table 1. The composition of the bacterial cell wall not only affects Gram-staining but also protects the bacteria from various medicines and immune system responses. The cell wall composition differs between Gram-negative and Gram-positive bacteria (Figure 1) and in both cases it forms a tight barrier against the outside environment. Few enzymes in the human body are capable of destroying cell wall which results in the death of the bacterium. Some microbial secondary metabolites, for example, penicillin prevents the formation of a water-tight Gram-positive cell wall, which causes the destruction of certain bacteria. The Gram-negative bacteria contain complex cell wall structure which protects the bacterium from immune system attack and prevents many antibiotics from working. Penicillin, a well known fungal secondary metabolite, declared the beginning of the antibiotics. Later, antibiotics are considered as *magic bullet* to cure various diseases. By the mid-20th century, the golden age of antibiotics provided a bountiful arsenal of tools for the treatment of bacterial infections. Extensive investigations were performed on hundreds of microbial metabolites to control bacterial diseases. Conversely, as

this era came to a close, widespread resistance to antibiotics by pathogenic bacteria turned into a serious health problem for community; organisms that were once sensitive to these drugs now well survive under the same treatment. The Infectious Diseases Society of America documented those antibiotic-resistant pathogens that are most commonly encountered in the clinic as the ESKAPE pathogens [1, 2].

1.1.1. ESKAPE pathogens

ESKAPE pathogen incorporates both Gram-positive and Gram-negative organisms and includes *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species. These pathogens pose as a potent threat for the patients as they are responsible for the majority of hospital-acquired infections. Furthermore, their pan-resistance for widespread antibiotic therapies has required the use of previously retired techniques having high toxicity [3, 4]. Since decade, advance molecular biology techniques have deciphered to thorough information about individual resistance mechanisms in all these pathogens. Still the data showing the interplay between resistance mechanism and bacteria is not sufficient to understand how the bacteria acquire these mechanisms. In addition, data on the impact of clinical interventions to decrease the prevalence of resistance are also lacking.

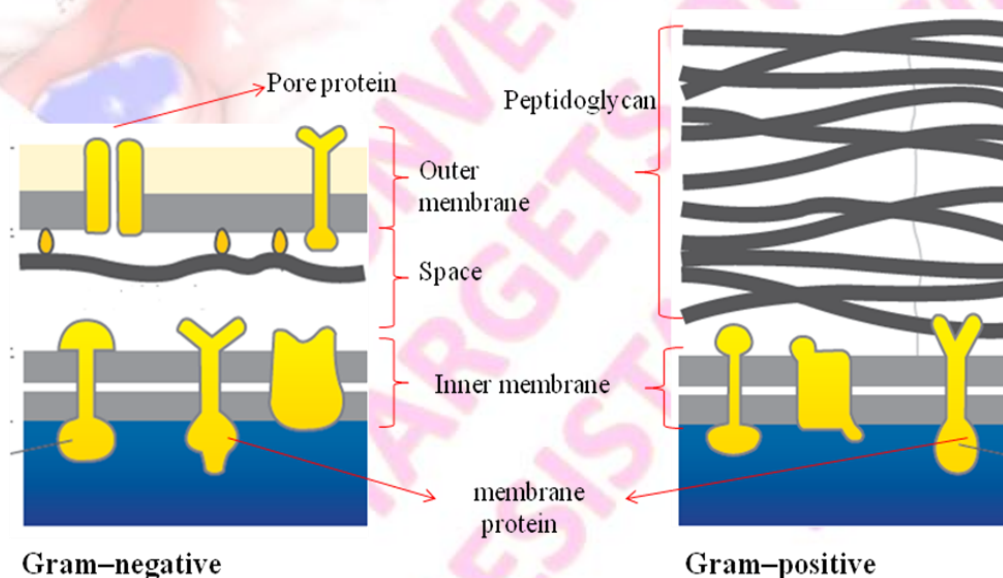


Figure 1. Bacterial cell wall, Gram-negative cell wall is more complex, with an outer membrane, a space, and a layer of peptidoglycan. The Gram-positive cell wall is bilayered with peptidoglycan as a major component.

The difficulty in identifying novel antimicrobial agents with reliable activity against these pathogens, as well as careful studies to identify optimal strategies for infection control and antimicrobial use are lacking. These ESKAPE pathogens generally share some common mechanisms to escape the effect of antibiotics *via* different routes. Herein we discuss the properties of these ESKAPE pathogens in brief.

1.1.1.1. *Enterococcus faecium*

The members of the genus *Enterococcus* are Gram-positive cocci and nonhemolytic, the well known of this genus is *E. faecium*, which is found as the commensal in the intestine of human, but it turns its nature from commensal to pathogenic and causes diseases such as neonatal meningitis or endocarditis. In the 1970's and 1980's the antibiotic resistant *Enterococci* was emerged as a leading cause for various infections such as urinary tract, bloodstream, and surgical wounds. The members *Enterococcus faecalis* and *Enterococcus faecium* are more widespread species in the hospital borne diseases, comprising approximately 90% of clinical isolates. The bacterium *E. faecium* is known to be responsible for the 10,000 to 25,000 deaths per year in the USA. The resistance for the most of the antibiotics is common among *Enterococci*, especially vancomycin-resistant *E. faecium* are of great concern as it found to contribute up to 80% of some hospital isolates. The possible treatment for the *E. faecium* infections are intake of linezolid, daptomycin, quinupristin, dalfopristin and sultamicillin. However, the resistance for the antimicrobial agents used to cure *E. faecium* infections has already been depicted.

1.1.1.2. *Enterobacter* species

The members of the *Enterobacter* species are Gram-negative bacilli usually inhabitant of the gastrointestinal tract. Examples of such organisms include *E. coli*, *E. cloacae*, *Proteus mirabilis*, and *Citrobacter freundii*. The members of *Enterobacteriaceae* are responsible for approximately one third of all cases of ICU-acquired pneumonia, urinary tract infection, and 10–15% of the ICU-acquired bloodstream infections. Treatment options for the *Enterobacteriaceae* are the use of intake of β -lactam antibiotics, β -lactam antibiotics combined with β -lactamase inhibitors, quinolones, TMP-SMX, aminoglycosides, and tigecycline [5–7].

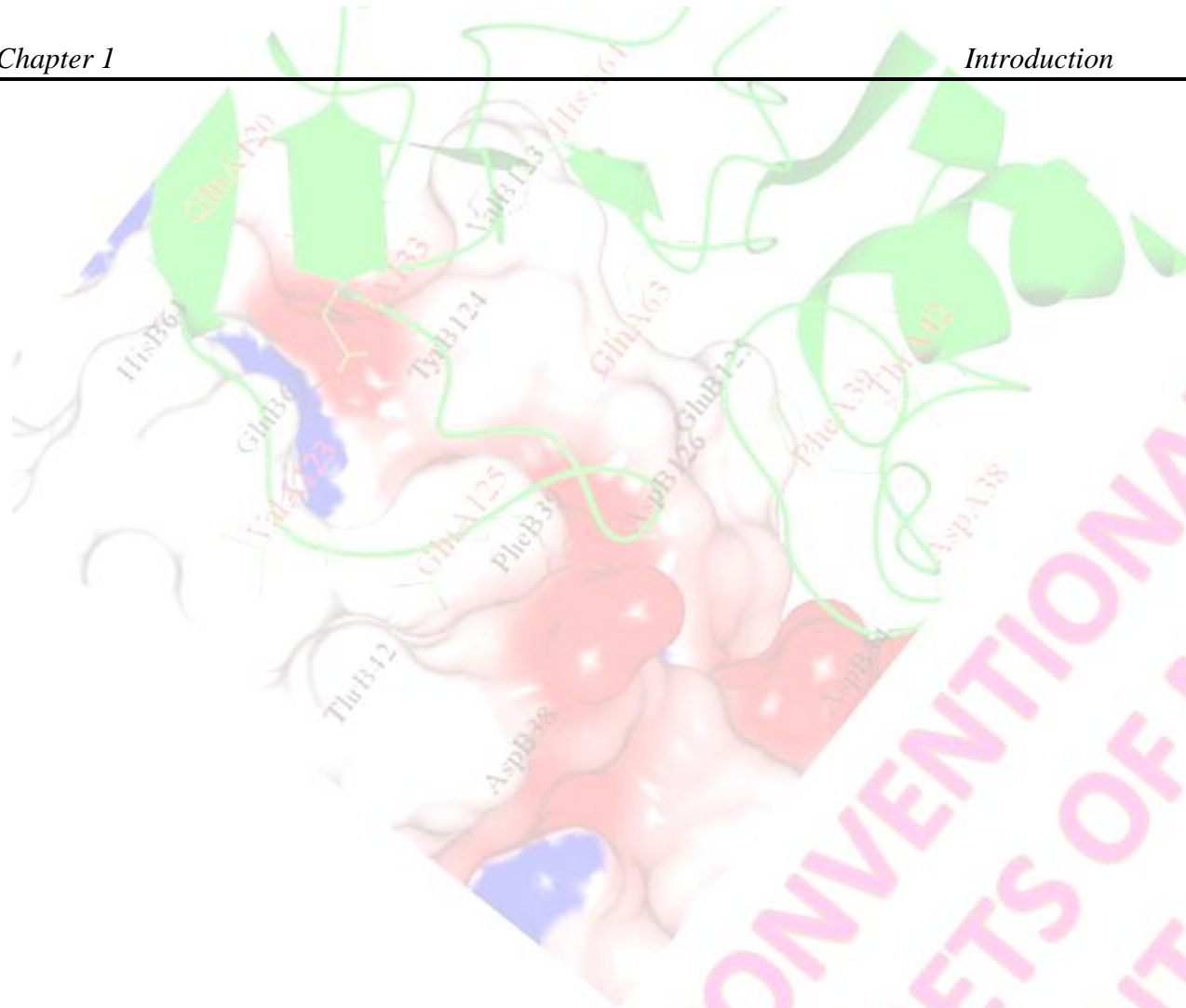


Table 1 . Examples of Gram-positive and Gram-negative bacteria and the diseases they cause in the humans.

UNCONVENTIONAL PROTEIN
TARGETS OF MULTI DRUG
RESISTANT ACINETOBACTER
JYOTI S. TOMAR
IITR

1.1.1.3. *Klebsiella pneumonia*

Klebsiella bacteria show resistance for a wide range of antibiotics. It is a Gram-negative nonmotile, encapsulated, lactose-fermenting, facultative anaerobic, bacterium. *Klebsiella pneumoniae* is a member of family *Enterobacteriaceae*, genus *Klebsiella*. It belongs to the normal flora of the human mouth and intestine. *K. pneumoniae* is very common and clinically more significant in comparison to other species of *Klebsiella*. Infections caused by the *K. pneumoniae* are usually community-acquired and primarily it infects the person with weakened immune system from improper diet. The rising rate and frequency of antibiotic resistance in strains is becoming more problematic for clinicians. *K. pneumoniae* is an opportunistic pathogen and confers resistance for most of the available antibiotics. *Klebsiella* spp. causes nosocomial, pneumonia, urinary tract, bacteraemia, neonatal sepsis and wound infections. *K. pneumoniae* consist of both carbapenemases and extended-spectrum beta-lactamases (ESBL), which are capable of hydrolyzing newer carbapenem drugs. ESBL-producing *K. pneumoniae* is resistant to other antibiotics, including fluoroquinolones, aminoglycosides, trimethoprim, and sulfamethoxazoles and more often *K. pneumoniae* shows resistance to cephalosporins and carbapenems. Recently, the emergence of MDR and pan-drug resistant (PDR) *K. pneumonia* strains have been shown to manifest the well known three mechanisms of drug resistance namely – the acquisition of novel antibiotic catalytic genes, mutations of antibiotic targets, and differential expression of efflux pumps. The best possible treatment for the patient infected with KPC-producing organisms is yet to be defined. The antimicrobial options are generally limited to polymyxins, tigecycline, and less frequently aminoglycoside antibiotics [8–10].

1.1.1.4. *Staphylococcus aureus*

Staphylococcus aureus is a Gram-positive bacterium frequently found in respiratory tract and on the skin. Normally *S. aureus* is non-pathogenic and silently stays as our natural flora but emergence of antibiotic-resistance phenotypes of *S. aureus* (eg., MRSA) makes it one of the most intractable pathogenic bacteria in the history of antibiotic chemotherapy. It conquered practically all the antibiotics that have been developed since 1940s. The pathogenic stain is found to be causing serious infections such as osteomyelitis, endocarditis, meningitis, bacteremia, septic arthritis, and nosocomial pneumonia. Treatment options for the *S. aureus*

infections include the use of intake of linezolid, quinupristin-dalfopristin, daptomycin, and tigecycline. Except the linezolid, these antimicrobial agents require intravenous administration [11].

1.1.1.5. *Pseudomonas aeruginosa*

P. aeruginosa is a Gram-negative, aerobic, coccobacillus bacterium. It is an opportunistic pathogen of immuno-compromised individuals. *P. aeruginosa* is the most frequent and troublesome member of the Gram-negative bacilli. It is a cause of ventilator-associated pneumonia, bloodstream infection and cholangitis. Antibiotic resistance posed by *P. aeruginosa* is a serious problem and it displays a diverse array of antibiotic resistance mechanisms [12]. *P. aeruginosa* produces AmpC β -lactamase, which can hydrolyze anti-pseudomonal penicillins, aztreonam, and third-generation cephalosporins. Another important mechanism associated with multidrug resistance is efflux pumps; they impart resistance to quinolones, cephalosporins, antipseudomonal penicillins and sometimes to aminoglycosides. The drugs tigecycline and quinolone are found to be ineffective against *P. aeruginosa* due to the presence of efflux pumps and mutations in the topoisomerases II and IV. Doripenem and biapenem are new carbapenems and found to exhibit good activity against the bacterium *P. aeruginosa*. In addition to these drugs, the polymyxins which include colistin or polymyxin E and polymyxin B, are found to be the consistently effective agents against highly drug-resistant Gram-negative bacteria, *P. aeruginosa*.

1.1.1.6. *Acinetobacter* species

Acinetobacter baumannii is a rod shaped Gram-negative bacterium; it is an opportunistic pathogen in humans and affects people with compromised immune systems. The ecology of bacteria belonging to the genus *Acinetobacter* is very diverse, and the members belonging to this genus have been recovered from soil, surface water, vegetables, animals, human body lice, and human hands. The widely accepted fact about the *A. baumannii* is its terrific ability of colonization, but not the infection. The growing numbers of *A. baumannii* strains and their resistance is considered as a major issue for health concern. It is still mysterious or unsolved question that which factors of *A. baumannii* decide its clinical success. In the current chapter, we discuss about *Acinetobacter* species, including their taxonomy, epidemiology, antibiotic resistance, clinical features and pathogenesis.

Different terms like multidrug resistant (MDR), extensive drug resistant (XDR) and pan drug resistant (PDR) have been used with varied definitions to describe the extent of antimicrobial resistance among spp. of *Acinetobacter*. Although till date, there are no accepted definitions for the extent of resistance in the bacteria and arbitrarily used terms have thus caused great confusion making it difficult for the available literature to be analyzed. Similar to the definition given by Falagas and co-workers [13], in this thesis, MDR *Acinetobacter* spp. is defined as the isolate resistant to at least three classes of antimicrobial agents — all penicillins and cephalosporins (including inhibitor combinations), Fluoroquinolones, and aminoglycosides. The XDR *Acinetobacter* spp. will be the *Acinetobacter* spp. isolate that is resistant to the three classes of antimicrobials MDR described above and along with resistance to carbapenems. Finally, PDR *Acinetobacter* spp. is the XDR *Acinetobacter* spp. that is resistant to polymyxins and tigecycline (Figure 2).

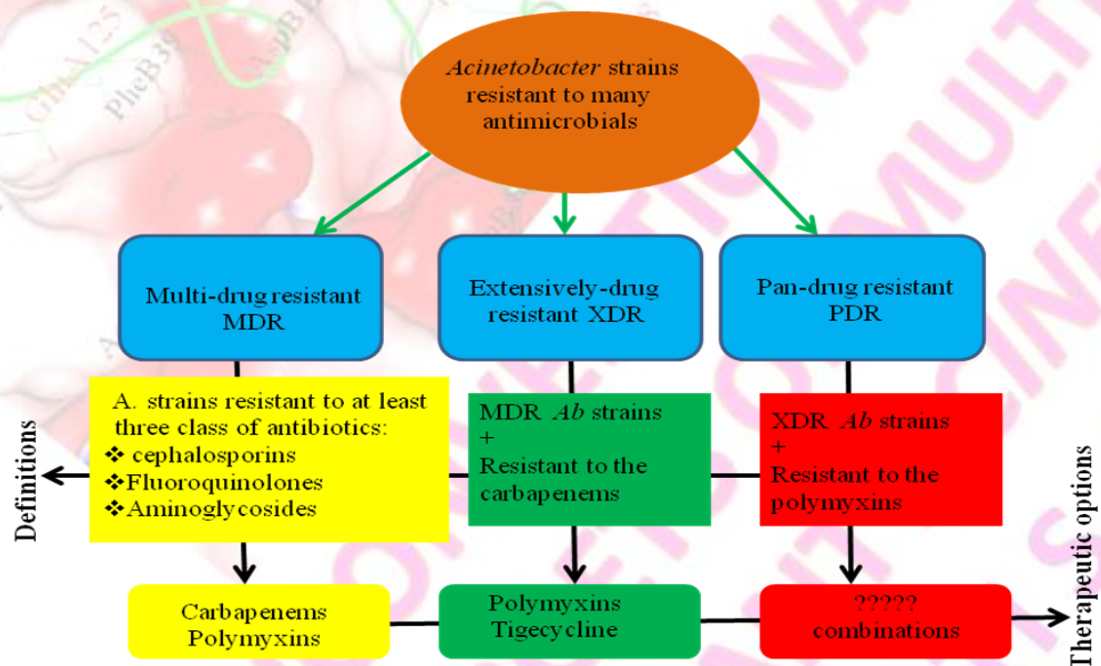


Figure 2. Definition of antibiotic resistant *Acinetobacter* strains along with therapeutic options.

Treatment of such multiple drug resistant *Acinetobacter* spp. is an incomprehensible scientific concern. The incidence of various infections caused by these organisms is difficult to resolve due to multiple risk factors associated with their antibiotic resistance. At the same time, a significant increase in the number as well as type of infections has been observed throughout the globe. Herein, we shed some light on the lifestyle and virulence factors of *Acinetobacter*.

1.1.2. Habitat of *Acinetobacter*

Acinetobacter spp. are widely distributed in soil and water and are found to grow at wide ranges of temperature, pH and other environmental conditions [14, 15]. In human, *Acinetobacter* is most commonly found to inhabit on skin, respiratory secretions, oropharynx secretions, and intravenous solutions. *A. baumannii*, the most important nosocomial *Acinetobacter* spp., has been rarely found on human skin although in hospitalized patients, the skin carriage rate of *Acinetobacter* has been found to be very high. The main reservoir for this bacterium is not well understood, whereas common sources of *A. baumannii* colonization or infections in the hospitalized patients are briefly explained in Table 2. *Acinetobacter* has capacity to survive on both dry and moist surfaces for long periods [16]. In the laboratory, the *A. baumannii* can be grown on agar plate, after 12 hour of incubation, it forms convex, circular, smooth and slightly opaque colonies, which are ~2.0 mm in diameter and pale yellow to greyish white in colour.

Sources of <i>Acinetobacter</i> species in hospital environment causing infections	Risk factor for the infections of <i>Acinetobacter</i>
Hands of the hospital staff	Underlying severity of illness
Respiratory therapy equipment	Exposure to antimicrobial agents like, carbapenems, colistin, etc
Food, Tap water	Exposure to an intensive care unit (ICU)
Infusion pumps	Colonization pressure
Mattresses, pillows etc in vicinity of patient	Invasive procedures, recent surgery
Hospital sink taps	Receipt of mechanical ventilation
Hospital floor	Prolonged length of hospital stay

Table 2. Sources and risk factors of infection with multidrug-resistant *Acinetobacter* species in a hospital environment.

Many case-control studies, with regard to *A. baumannii*, have revealed that prior exposure to broad-spectrum antibiotics is the most common risk factor for the infections. For example, the

patients with cystic fibrosis are highly susceptible for lung infections caused by the MDR *A. baumannii* because such patients are violently treated with a variety of antibiotics during the treatment of the same disease. The most commonly used medicines in the treatment of cystic fibrosis include carbapenems and third generation cephalosporins, followed by fluoroquinolones, aminoglycosides, and metronidazole. The other risk factors include hospital settings such as mechanical ventilation, a stay in an ICU, recent surgery, illness severity and other medical associated invasive procedures [17–21]. During the outbreaks, excessive contamination of the environment, including respirators and air samplers in the vicinity of the infected or colonized patients and from mattresses, pillows and blankets in the immediate vicinity of infected patients have been documented. Studies performed to understand the causative factors for *A. baumannii* outbreaks have revealed environmental contamination as an important source. The community-acquired habitat of *A. baumannii* infection includes its reservoirs at the outside of hospital, and many potential interactions with human infections. The community-acquired *A. baumannii* diseases include, community-acquired pneumonia, infections in survivors from natural disasters. The well known community-acquired *A. baumannii* infections in survivors from natural disasters are infected war wounds in troops from Iraq and Afghanistan. The emergence of community-acquired *A. baumannii* infections is mainly associated with interactions between animals, environment, and humans that are considered to be potentially involved in the emergence of some infectious diseases. Though the existence of extra-hospital reservoirs of *A. baumannii* has long been disputed but recent advancement in the molecular methods has allowed the demonstration of the actual presence of this organism in various environmental locations.

1.1.3. History and geographical distribution of *Acinetobacter baumannii*

Acinetobacter baumannii was first isolated in 1911 from a soil sample by Beijerinck. The first *Acinetobacter* spp. was thought to be non-virulent saprophytes [22]. *Acinetobacter baumannii* was taxonomically classified in 1986 by Bouvet and co-workers [23]. First carbapenem-resistant *A. baumannii* isolates were discovered in the year 1991 [24], which also document the first outbreak of carbapenem-resistant *A. baumannii* in the USA [25]. Carbapenem-resistant *A. baumannii* isolates were isolated from a leukemia patient and these isolates were found to be resistant to almost all antibiotics *eg.*, cephalosporins, aztreonam, aminoglycosides and ciprofloxacin and were therefore named pan-drug resistant *A. baumannii* (PDR-Ab) [26].

The rise in the number of multidrug resistant *A. baumannii* (MDR-Ab) strains has been due to the extensive use of antimicrobial chemotherapy against bacterial infection. Various species of genus *Acinetobacter* have been detected in Australia, Brazil, China, Germany, India, South Korea, United Kingdom and United States. The antibiotic susceptibility data of *Acinetobacter* from different geographical regions revealed that the resistance for imipenem was in the range of 0–40%. In India from Vellore, a prevalence of 14% carbapenem resistant in *Acinetobacter* spp., was reported and in Delhi, India (2006), the prevalence of carbapenem resistance in *Acinetobacter* spp., isolated from different clinical samples, was found to be almost 35% [27]. The above mentioned data suggests that the rate of resistance development can be slowed down if the antibiotic therapy is done, after checking the *in vitro* susceptibility profile of the organism for the same antibiotics.

1.1.4. Classification of *Acinetobacter baumannii*

Acinetobacter are grouped into three main complexes: i) *Acinetobacter calcoaceticus-baumannii* complex, which is glucose oxidizing and non-haemolytic; ii) *Acinetobacter lwoffii*, which are glucose negative and non-haemolytic and iii) *Acinetobacter haemolyticus*, which is haemolytic [23, 27]. The full classification of *A. baumannii* is listed in Table 3. In 2008, Euzéby provided the nomenclature of *Acinetobacter baumannii*, where he has recognized 21 genomic species of the *Acinetobacter* genus are given below and 14 more recognized genomic species are still unnamed. The 21 genomic species of *Acinetobacter* are *A. baumannii*, *A. baylyi*, *A. beijerinckii*, *A. bouvetii*, *A. calcoaceticus*, *A. gerneri*, *A. grimontii*, *A. gyllenbergii*, *A. haemolyticus*, *A. johnsonii*, *A. junii*, *A. lwoffii*, *A. parvus*, *A. radioresistens*, *A. schindleri*, *A. soli*, *A. tandoii*, *A. tjernbergiae*, *A. townneri*, *A. ursingii*, and *A. venetianus*.

Domain	<i>Bacteria</i>
Phylum	<i>Proteobacteria</i>
Class	<i>Gammaproteobacteria</i>
Order	<i>Pseudomonadales</i>
Family	<i>Moraxellaceae</i>
Genus	<i>Acinetobacter</i>
Species	<i>Baumannii</i>

Table 3. Classification of *Acinetobacter baumannii*.

Acinetobacter baumannii are Gram-negative, oxidase-negative, non-motile, non-fermentative, aerobic *coccobacillus*. *Acinetobacter baumannii* are ubiquitous in nature and commonly found within the ICU of hospital *i.e.*, the environment causing a variety of opportunistic nosocomial infections. *Acinetobacter* spp., are not able to show any morphology change in Gram-stained human clinical specimens due to which the *Acinetobacter* cannot be differentiated from other causes of nosocomial infections. Infection caused by *A. baumannii* is difficult to cure due to the Gram-negative nature of the cell wall as the outer wall provides a barrier so that the antimicrobial agent is unable to enter the bacterial cell. *A. baumannii* is an opportunistic pathogen and is able to resist desiccation for long period of time [28, 29]. The bacterium is also able to survive on inanimate surfaces for months to years [30] and is the common cause of nosocomial infections associated with the mechanical devices. This bacterium is able to resist both physical and chemical disinfections due to its capacity of biofilms formation [31, 32]. *A. baumannii* infections include diverse contaminated surfaces, which work as an important source of secondary infection. Infected or colonized patients are important reservoirs of *A. baumannii*. This bacterium is passed from patient to patient *via* direct and indirect contact.

Generally endemic strains are found to coexist with one or more than one epidemic *Acinetobacter* species clones, making it tough to detect and control transmission. *Acinetobacter* has capacity to form biofilms in various environments and on the surface of various implants. The ease to survive in the hospital environment and increasing antibiotic resistance in *acinetobacter* is also contributed by its ability to form biofilms. Due to this reason, the antibiotic for which it is having *in vitro* susceptibility will also be ineffective in treating the infection.

1.1.5. Pathogenesis of *Acinetobacter baumannii*

The members of the genus *Acinetobacter* are believed to have restricted pathogenic potential against healthy people, while in immune-compromised patients, some species, for example, *A. baumannii* can cause severe, life threatening infections. The infections caused by *Acinetobacter baumannii* are associated with organs like lungs, cerebrospinal fluid, peritoneal fluid and the urinary tract and predominantly occur in the patients suffering from other disease [33]. Disease symptoms include nosocomial pneumonia, urethritis, catheter associated *bacteriuria*, wound infections, and peritoneal dialysis associated peritonitis. *A. baumannii*

infections very rarely include meningitis, urinary tract infections, endocarditis and cholangitis. Airborne spread of *A. baumannii* was first time documented by the Allen and Green in 1987 [34]. The mortality rate associated with the *A. baumannii* infection is high, for example, mortality rate is about 23%, for the hospitalized patients and a 43% mortality rate for the patients in intensive care [35]. The antimicrobial availability task force (AATF) of the infectious disease society of America designated the *A. baumannii* as one of the principally problematic pathogens and there is a desperate necessity for new drug development to cure the infection caused by this bacterium. However it is not easy to distinguish between colonization and infection regarding *A. baumannii*. Options for treating the community acquired *A. baumannii* are carbapenems, cefopime, cefepime, ceftazidime, aminoglycosides and fluoroquinolones [36]. *Acinetobacter* strains possess only some virulence factors however some strains of it have virulence factors allied with invasiveness, transmissibility or the higher ability to colonize in immune-compromised patients. A new strain, OXA-23 clone II, identified in a military hospital, was found to be a highly virulent strain and extremely difficult to eliminate from medical facilities. A large portion of the *A. baumannii* genome is made up of virulence islands having a large number of genes dedicated to pathogenesis [37]. *A. baumannii* and *Acinetobacter* DNA group13TU is found to be responsible for the majority of *Acinetobacter* hospital outbreaks. Till date, *A. baumannii* has caused numerous global outbreaks and exhibited constantly increasing rates of resistance. Recently, emergence of imipenem-resistant strains of *A. baumannii* has been predicted in the hospital outbreaks [38]. Pathogenesis of *A. baumannii* is considered as the outcome of its capability of resistance for the multiple antibiotics. Now the question is what are the factors or traits which help in acquiring antibiotic resistance and how antibiotic resistance makes it as one of the serious life threatening problems. To answer this question, first of all we need to know the physical and chemical properties of the antibiotics. The mechanisms associated with the resistance development are described below:

1.1.6. Antibiotics

Antibiotics are natural microbial metabolites having low-molecular-weight and have capacity to inhibit the growth of other microorganism at optimum concentrations. Inhibitions may be permanent or temporary and the permanent inhibition caused by antibiotic action is called as bactericidal, whereas in the later case the inhibition is lost when antibiotic is removed from

the medium called as bacteriostatic action of that antibiotic. The products obtained by chemical modification of natural antibiotics or from other products of microbial metabolism are called semi-synthetic antibiotics. Antibiotics cannot be classified as a homogeneous category of chemical substances as they have very diverse chemical nature and only common property among them is that they are organic solids. Next thing which need to be known is the origin of these compounds, but answer for this is not straightforward because these compounds are produced by an array of widely diverse microorganisms. More than 50% of antibiotics are produced by members of order Actinomycetales, and genus streptomycetes. Biosynthesis of antibiotics can be grouped into a few fundamental biosynthetic pathways and on the basis of biosynthetic pathways the antibiotics are further categorized as (i) peptide antibiotics, (ii) terpenoid antibiotics, and (iii) aminoglycoside antibiotics. Analogous to primary metabolites, antibiotics derived by the polymerization for example peptide antibiotic, antibiotics derived from the acetate and propionate units and fatty acid synthesis pathway, terpenoid antibiotics, and aminoglycoside antibiotics which are derived by condensation of sugar or amino sugars and cyclic amino alcohol. The spectrum of antibiotics activity can be grouped in two classes, the antibiotics effective against Gram-positive bacteria only called as *narrow spectrum of activity*. On contrary, the antibiotics effective against both Gram-positive and Gram-negative bacteria are said to have *broad spectrum of activity* and this activity spectrum depends upon the structure of their cell walls.

In a given population, a few microorganisms that are not inhibited by the antibiotics at the same concentration may be present which are sufficient to inhibit the majority of cells and these microbes are known as the mutants. The frequency of these mutants generally ranges between one mutant in every 10^7 to 10^{10} sensitive cells. However, when a microbial population having few mutants is exposed to inhibitory concentration of antibiotic then growth of sensitive cells is blocked while mutants continue to multiply and finally the whole population will become resistant, which indicates that antibiotic exposure effectively act for the selective development of resistant population. In the absence of same antibiotic, the formation of sensitive cells may occur with a high frequency by the back mutation, and in long term the population reverts back to sensitivity. The antibiotics block the growth of sensitive microbes by the inhibition of macromolecules, which particularly play role in cell wall, protein or RNA synthesis, DNA replication or membrane function.

1.1.6.1. Foundation of the antibiotic era

In the early 1940's industrialization of penicillin worked as the inauguration step in the successful era of antibiotic innovation that lasted for over two decades (Figure 3). In this era, the discovery of many new antibiotics classes, tetracyclines, macrolides, aminoglycosides, cephalosporins, chloramphenicol, glycopeptides and rifamycins, which are nowadays used clinically, was done [39]. However, in the 1960's, when a first case of resistance in microbes towards these antibiotics was observed, the golden age came to an end [40]. As resistance development limits the lifespan of antibiotics due to which a constant requirement of introduction of new compounds is necessary. After this era, only two structurally different antibiotics have been released into the clinic, namely, linezolid and daptomycin. The invention of new antimicrobial drug is exceptionally difficult principally due to complex cell wall structure which limits the penetration capacity of molecules into bacterial cells. On the other hand, natural antibiotics have although evolved to violate the penetration barriers of target bacteria. Most of the antibiotics were discovered by screening cultivable soil microorganisms and excess mining of this natural resource brought an end to the initial era of antibiotic discovery. In the last decades, the battle between bacteria and science has gone both ways but,

Year of discovery	Antibiotic(s)
1928	Penicillin
1932	Sulfonamides
1943	Streptomycin
1946	Chloramphenicol
1948	Cephalosporins
1952	Erythromycin and Isoniazid
1957	Vancomycin
1961	Trimethoprim
1976	Carbapenems
1979	Monobactams
1987	Lipopeptides

Table 4. Discovery of antibiotics.

as we are approaching the end of the second decade of the 21st century, undeniably the bacteria are winning. Resistance in bacterial strains is becoming common phenomenon, which makes them responsible for the most of hospital-acquired infections observed in the best medical centers and hospitals of all over the world. Pharmaceutical companies involved in antibiotic discovery by derivation of existing antibiotics classes are also facing failure in fighting the battle against resistance [41]. Researchers are still trying to develop new types of antibacterial compounds but, as fast as they are produced, bacteria build up strategies to escape their effects. The well known antibiotics with year of their discovery, widely used by clinician so far are given in Table 4.

1.1.6.2. Classification of antibiotics

Various schemes for the classification of antibiotics have been proposed, but none of them are universally adapted. Antibiotics those have a common basic chemical structure are grouped into one class are classified as described below:

1.1.6.2.1. β -Lactam antibiotics

The *penicillins* are the first discovered antibiotics and still work as best option for treatment of many infections. The penicillins along with *cephalosporins* form the group of β -lactam antibiotics, and are named so because of the presence of four-atom cyclic amide in their structure chemically pronounced as β -lactam. This group of antibiotics causes irreversible damage to the bacterial cell wall by inhibiting the synthesis of peptidoglycan, basic component of the bacterial cell wall. They are bactericidal antibiotics and ineffective against fungi and mycoplasma as both of these organisms do not have cell wall. These molecules are more successful against the Gram-positive bacteria and toxicity of penicillins and cephalosporins is very low.

1.1.6.2.2. Tetracyclines

This family of antibiotics is known for its broad spectrum of action and its great therapeutic effectiveness. Biologically these antibiotics are synthesized by cyclization of a chain obtained by condensation of acetate and malonate units. These molecules are named as tetracyclines because its chemical structure consists of four rings condensed linearly. They act on ribosomal sub-unit to inhibit bacterial protein synthesis and as their effect are reversible, and they are

thus grouped as bacteriostatic agent. Activity spectrum of these compounds is broad including Gram-positive, Gram-negative, rickettsiae, chlamydiae, and some protozoa.

1.1.6.2.3. Aminoglycosides

This class of antibiotics is characterized by the presence of cyclic amino alcohol and comprise of large number of substances. *Streptomycin* is the first discovered aminoglycoside, and other well known examples of this class are *kanamycin*, *gentamycin*, *amikacin*, and *netilmicin*. The aminoglycoside antibiotics inhibit protein synthesis irreversibly by interacting with the ribosome subunits and exhibit bactericidal effect. These compounds are actively used against Gram-negative bacteria. These compounds are highly soluble in water due to the presence many hydroxyl functionalities in their structure and as a result, these are not suitable for the oral uptake. Upon oral uptake, major adverse toxic effects, including nephro-toxicity and ototoxicity, are detected.

1.1.6.2.4. Macrolides

The chemical structure of these antibiotics is characterized by a ring consisting of no fewer than 12 carbon atoms and closed by a lactone functionality. These compounds are biologically synthesized by the condensation of a number of propionate and acetate units. Further they can be divided in two homogeneous classes; the compound in the first subclass are antifungal and antiprotozoal macrolides, characterized by the lactone rings of about 30 carbon atoms with hydroxyl groups and a series of conjugated double bonds due to which they are also known as *polyenes*. They induce distortions in the cell membranes by interfering with the sterols; high toxicity of these compounds limits their use and well known member of this group is *amphotericin B*. The compounds in the second class are antibacterial macrolides, which are characterized by lactone rings of 14 or 16 carbon atoms. They inhibit the protein synthesis reversibly by interacting with ribosome and show bacteriostatic action. They act against Gram-positive bacteria and mycoplasma. Examples of this group are *erythromycin*, *spiramycin* and *leucomycins*.

1.1.6.2.5. Ansamycins

This family of antibiotics, introduced in 1960's, consists of molecules which are made up of an aliphatic chain that connects two opposite points of aromatic ring. Their biosynthesis

resembles with the biosynthesis of the macrolides. They are further divided in two subclasses viz., benzenes and naphthalenes. Naphthalenes inhibit selectively enzyme polymerase and show antibacterial action, while benzenes are not selective for its action and generally studied for antitumor property. The well known example of the subclass naphthalenes is rifamycin which shows broad spectrum of activity.

1.1.6.2.6. Peptide and glycopeptides antibiotics

The chemical structures of these antibiotics consist of a chain of amino acids but their biosynthesis pathway is not similar to protein synthesis. First peptide antibiotics discovered is gramicidin, which interferes with the membrane function. This group of antibiotics exhibits high toxicity due to which generally these are not used by the medical practitioners and among these, only polymyxins are used systematically. The polymyxins are known to disrupt the membrane function, due to which they are very effective against the Gram-negative bacteria. Glycopeptides are made up of linear heptapeptides and inhibit the bacterial cell wall synthesis by making complex with intermediate moiety D-alanyl-D-alanyl. The important member of this class is *vancomycin*, and it is active against Gram-positive bacteria *staphylococci* resistant to β -lactams.

Some antibiotics which cannot be classified into any of the families mentioned above, are described here: (i) *Chloramphenicol*, is bacteriostatic and inhibit the protein synthesis. It is very effective against Gram-negative bacteria. (ii) *Novobiocin*, active against Gram-positive bacteria, interferes with the DNA by inhibiting DNA gyrase. (iii) *Fusidic acid*, steroid type structure, active against Gram-positive bacteria, it inhibits the elongation factor of protein synthesis.

1.1.6.3. Antibiotic resistance - what, why and how?

Before coming to the other aspects of antibiotics here we discuss what antibiotic resistance is. The bacteria having ability to survive in the presence of drugs that would normally kill it is termed as antibiotic-resistant. Resistance for a particular antibiotic is measured by determining its minimum inhibitory concentration (MIC), which corresponds to the minimum drug concentration that can be used to cure the infection. The MIC of an antibiotic should be enough to cure disease as well as not too high which cannot be tolerated by patients. Susceptible bacteria have a MIC that falls within the standard dosing range of the antibiotic.

While resistant bacteria requires high doses of antibiotic that have not been revealed to be tolerable for the patients.

Why resistance for the antibiotics has evolved so widely? As in the mid of 20th century, antibiotics were considered as miracle drugs but nowadays antibiotics are renowned as a double-edged sword. In the mid of 20th century antibiotics provided abundant store of weapons to cure the bacterial infections due to which this era was considered as the golden age of antibiotics. Lately, due to pervasive development of antibiotic resistance by many of disease causing bacteria became a serious health care-associated problem.

The antibiotic-resistant bacteria *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* (ESKAPE), as documented by Infectious Diseases Society of America, are discussed in brief in the following section. These ESKAPE pathogens are capable of neutralizing the biocidal action of antibiotics and communally representing novel paradigms in pathogenesis, transmission and resistance. Before a decade concern of clinical microbiologists was centered on Gram-positive bacteria (meticillin-resistant *Staphylococcus aureus* and vancomycin resistant *Enterococcus* spp.) but now, rate of increase in the resistance in Gram-negative bacteria is faster than in Gram-positive bacteria.

Another key concern is that there are only fewer new and developmental antibiotics active against Gram-negative bacteria. As per present scenario, the drug development agenda will not be enough to provide therapeutic cover in coming 10–20 years. Majority of the antibiotics used today to cure the diseases caused by the bacteria are originated from actinomycetales, mostly streptomycetes, and were isolated during the golden age of antibiotics discovery. Natural products have proven to be highly efficient for the treatment of bacterial infections and the variety of drugs based on natural products is enormous. There are drugs with broad and narrow spectra for oral, topical or parenteral administration and with activities against almost all known pathogens. Most natural product-based candidates, currently under development, are either new or improved versions of old drugs. Although chemical modification of the existing drugs has been demonstrated to be the most efficient way to develop new drugs active against the resistant strains, these new drugs are doomed to suffer from resistance development as well. Therefore, the development of antibacterial drugs with completely new modes of action is needed. The screenings of libraries containing between 260,000 and 530,000 molecules resulted in five candidates, none of which have subsequently

passed clinical trials to become licensed. This study reveals the complexity in finding novel antibacterial drugs. In the present scenario of the fight against bacteria infectious, researchers are slowly losing the battle, because of speedy antibiotics resistance development. By 2050 it is expected that drug-resistant infections will kill more people globally than cancer.

1.1.7. How do antibiotics resistance developed?

Before knowing about how antibiotic resistance is developed we need to know the mode of antibiotic action. Antibiotics can have different mechanisms of action that work against the infecting pathogen. The vast majority of antibiotics fall into the same class and work by preventing the cell wall of the bacteria from forming or inhibiting the creation of bacterial proteins needed for survival or reproduction. Antibiotics may also prevent bacterial genes from being copied, which keep the bacterium from reproducing. Figure 3 illustrates the general targets of antibiotics. The five major modes of antibiotic mechanisms of activity are described below with examples:

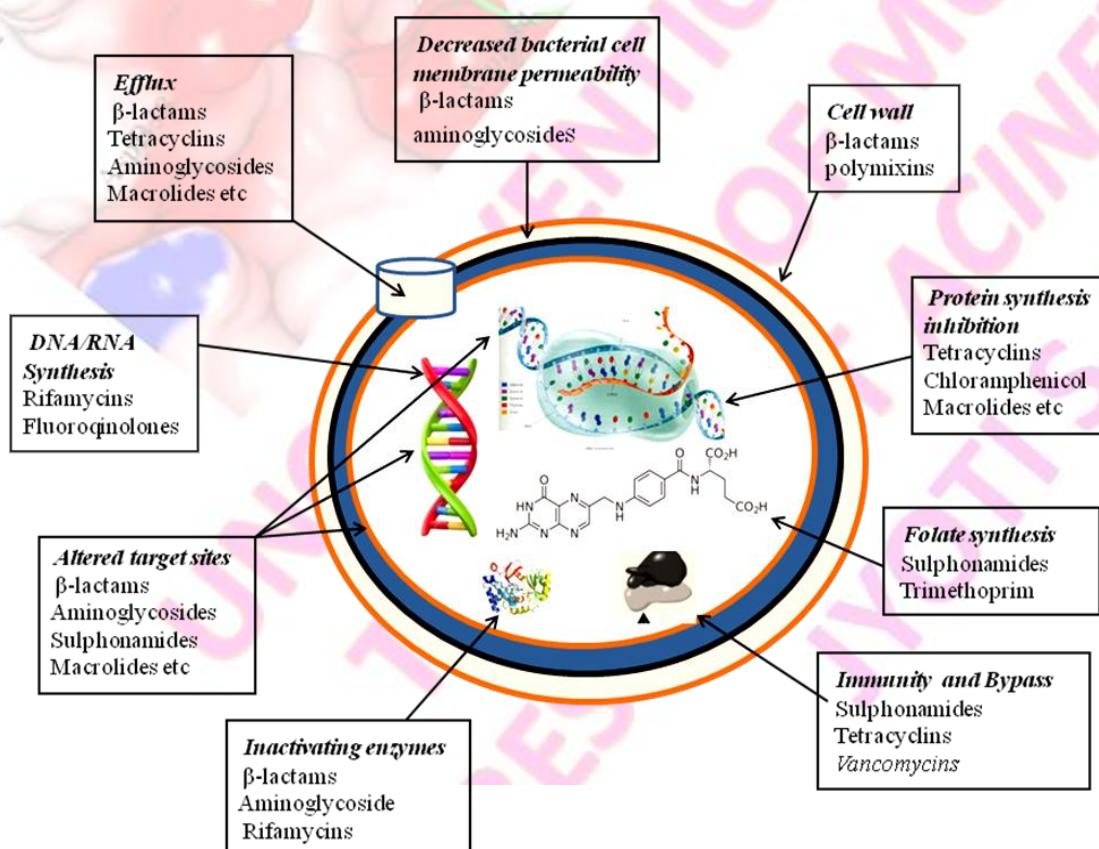


Figure 3. Main targets of antibiotics in the bacterial cell.

1.1.7.1. Interference with cell wall synthesis

The β -lactam antibiotics, like penicillin and cephalosporin hinder the enzyme action needed for the formation of peptidoglycan layer [42]. As peptidoglycan layer is essential component of cell wall hence this hindrance results in failure of cell wall synthesis.

1.1.7.2. Inhibition of protein synthesis

The antibiotics oxazolidinones interact with site A of the bacterial ribosome which results in the interference of the aminoacyl-tRNA placement. Tetracyclines obstruct the synthesis of protein by binding to 30S ribosomal subunit, thereby deteriorating the ribosome-tRNA interaction. Another example of antibiotic inhibiting protein synthesis is macrolides, which bind to 50S ribosomal subunit and stop the elongation of nascent polypeptide chains. Similar to macrolides, the chloramphenicol also binds to the 50S ribosomal subunit but instead of blocking elongation of peptide chains it inhibits the action of the enzyme peptidyl transferase. Lastly, the antibiotic class Aminoglycosides bind to the 30S ribosomal subunit and inhibits the initiation of protein synthesis in bacterial cell [43].

1.1.7.3. Interference with nucleic acid synthesis

Antibiotic rifampicin hinders the action of the enzyme DNA-directed RNA polymerase. The quinolones inhibit the type II topoisomerase, leading to inhibition of DNA synthesis during replication cycle and finally causing DNA double strand break [44].

1.1.7.4. Inhibition of a metabolic pathway

The well known examples of metabolic pathway inhibiting antibiotics are sulfonamides and trimethoprim which block the key steps of the folate synthesis. The folate works as a cofactor in the biosynthesis of DNA and RNA nucleotides.

1.1.7.5. Disorganizing of the cell membrane

The cytoplasmic membrane is considered as the primary site of action for many well known antibiotics. For example, polymyxins exert their inhibitory effect by increasing the bacterial membrane permeability and causing the leakage of bacterial content. Another well known class of antibiotics, lipopeptide daptomycins, display rapid bactericidal activity by binding to

the cytoplasmic membrane in a calcium-dependent manner, leading to an efflux of potassium from the bacterial cell and resulting in the cell death [45].

1.1.8. Bacterial strategies to develop antibiotics resistance

In 1946, Fleming stated that there is probably no chemotherapeutic drug with which in suitable circumstances, the bacteria cannot react and in some way acquiring fastness *i.e.*, resistance. This statement is widely accepted and nowadays it's well known that bacteria can adapt various strategies to evade the antibiotics including; efflux of drug from the cell, over expression of drug target, chemical modification of drug molecule and abating the target requirement **Figure 3**. For example, the well known drug chloramphenicol have become less effective and no longer considered as first line agent for any infection due to emergence of drug efflux systems in many bacteria. Rifampicin antibiotics known to inhibit RNA polymerase, has been shown to be less effective by the changes of its target.

The reduction in effectiveness of a drug such as antimicrobial on an anti-neoplastic in curing a disease or condition is known as antibiotic resistance. When an antibiotic is not capable to kill or inhibit a pathogen, then the term resistance is used. Normally, an organism showing resistant to more than one drug is said to be multidrug-resistant [46]. Bacterial strains may possess different types of resistant mechanisms which are shown in **Figure 5** and are explained as follows.

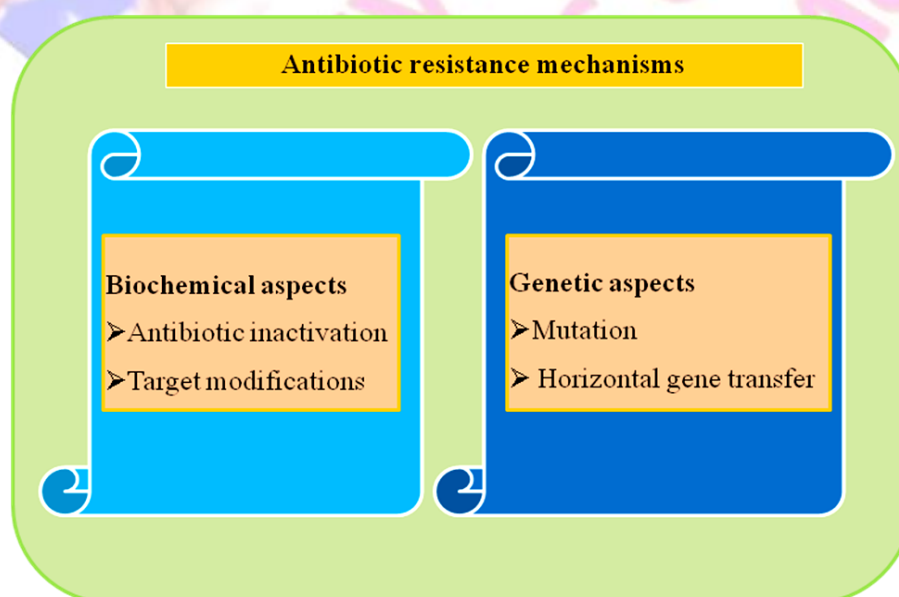


Figure 4. Biochemical and genetic aspects of antibiotic resistance.

1.1.8.1. Biochemical aspects of antibiotic resistance

1.1.8.1.1. Antibiotic inactivation

1.1.8.1.1.1 By redox process

In many pathogenic bacteria, oxidation or reduction of antibiotics was done to destroy their effects. Although this is not a frequent mechanism however, there are a few examples of this strategy [47]. For example, oxidation of tetracycline antibiotics by the TetX enzyme was performed to neutralize the effect. *Streptomyces virginiae*, a producer of the type A streptogramin antibiotic virginiamycin M1, protects itself from its own antibiotic by reducing a critical ketone group to an alcohol.

1.1.8.1.1.2. By hydrolysis

The antibiotics having chemical bonds in their structural moiety such as amides and esters are hydrolytically susceptible. Many enzymes are known to spoil antibiotic activity by cleaving these bonds. The well known class of enzyme is extended spectrum β -lactamases (ESBLs), which are known to confer the resistance to all penicillins, third generation cephalosporins (*e.g.*, ceftazidime, cefotaxime, and ceftriaxone) and aztreonam [48].

1.1.8.1.2. Antibiotic inactivation by group transfer

The most diverse family of resistant enzymes is the group of transferases. These enzymes are known to inactivate the antibiotics aminoglycosides, chloramphenicol, streptogramin, macrolides and rifampicin chemically by the addition or substitution of adenylyl, phosphoryl or acetyl groups. These groups are generally added to the edge of the antibiotic molecule to modify. Due to these modifications, the antibiotics cannot impair the binding to their target. The chemical strategies include *O*-acetylation and *N*-acetylation, [49, 50] *O*-phosphorylation, [51] *O*-nucleotidylation, [52] *O*-ribosylation, *O*-glycosylation, and thiol transfer. These modification processes are restricted to the cytoplasm and most of these modification strategies require a co-substrate such as ATP, NAD⁺, acetyl-CoA, UDP-glucose, or glutathione for their activity.

1.1.8.1.3. Antibiotic resistance via target modification

The second major resistance mechanism is the modification of antibiotic target site so that the antibiotic is impotent to bind properly. However, it is possible for mutational changes to occur

in the target that reduce the susceptibility for the inhibition while retaining cellular function [53].

1.1.8.2. Genetic aspects of antibiotic resistance

Generally bacteria develop antibiotic resistance through genetic changes that can occur when resistance genes transfer between different bacterial strains or between different bacterial species or when mutations take place and these mutations help to confer the resistance or the genes that help in survival under adverse conditions are picked up from the host or from the environment. Important genetic factors responsible for the development of the resistance are given in [Appendix I-1](#) and their mechanisms with examples are described below:

1.1.8.2.1. Mutations leading antibiotic resistance

The emergence of antibiotic-resistance *via.*, mutation is a very complex process, in which the genetics, the physiology, and the population behaviour, together with environmental parameters, play key roles. The mutation rate determined under conventional laboratory conditions most likely differs significantly from those of *in vivo i.e.*, at the site of infection. For example, resistance for antibiotics fluoroquinolones and rifampicin is acquired by mutations in the gene [54]. Furthermore, the efflux system of membranes and variation in the antibiotic uptake is also reported as a result of mutations [55].

1.1.8.2.2. Horizontal gene transfer leading Antibiotic resistance

Major cause of the antibiotics spreading is horizontal transfer of genetic material. Genes conferring antibiotic resistance are mostly transferred by three mechanisms namely conjugation, transformation and transduction. The well known example of this process is spreading of the β -lactamase enzymes. One of these, CTX-M-15, initially found in *E. coli* but now it reported in other members of Enterobacteriaceae [56, 57]. It is often located on highly mobile IncFII plasmids and associated with mobile genetic element IS26.

1.1.8.2.3. Genomic Duplications

Gene duplications is more common mechanism for drug resistance in the eukaryotic cells (certain mammalian cancers) as compare to prokaryotic cells. Gene duplications causes the over expression of the multidrug transporters and the drug targets [58]. In prokaryotic cells the

large tandem duplication is also reported for example in the *E.coli* *acrAB* locus in a mutant isolated in the presence of tetracycline, which was found to be associated with the over-expression of the AcrAB drug efflux pump, producing an MDR phenotype [59]. This mutant was found to be very unstable and again converted into the wild phenotype in absence of drug [60]. Another example of genomic amplification has also been found in *S. aureus*, which generates methicillin susceptibility [61].

1.1.9. Strategies to develop new antibacterial drugs

The efforts towards developing the effective and new antibiotics are required to prevent the constantly rising resistance in pathogenic bacteria. The one possible approach is to chemically modify the existing antibiotics to make them effective against the resistant species. This strategy is found to be milestone in case of β -lactam class of antibiotics, modification of these antibiotics resulted in the increased efficiency against resistant bacteria. But the resistance against these modified drugs (third generation β -lactams) has been emerged shortly. The second approach is to target resistance causing gene product. This approach has been implemented with mixed success. For example, clavulanic acid, an inhibitor of β -lactamases is effective when supplemented with amoxicillin, a β -lactam class of antibiotics. Although the above two strategies have been proved reasonably successful but rapid emergence of resistance made them less effective against resistant bacteria. The third approach is to develop the drugs which target the essential metabolic processes of bacteria. Till date most of the existing antibiotics target only three essential metabolic processes including DNA repair/replication pathway, protein biosynthesis pathway and peptidoglycan/cell wall biosynthesis pathway. But, due to selective pressure, drug resistant strains have been emerged against the members of all three classes. Hence, new pathways and targets are needed to be characterizing for the development of new class of potent antibiotics. Due to significant advancements in high throughput automated DNA sequencing, last decade witnessed the exponential rise in genome sequence data which can be utilized for the drug discovery and development process. The novel antibiotic development through this strategy involves many key steps including, the identification and structural/functional characterization of validated target, high-throughput screening against different compound libraries is used to narrow down the possible lead molecule. For the further validation and development of these leads toward

anti-bacterial, structure–activity relationship (SAR) study followed by testing in animal models and then in human is carried out **Figure 5** [62].

1.1.9.1. Present strategies with their shortcomings

As mentioned in the previous sections that unique feature present in *A. baumannii* is its notable array of acquired antibiotic resistance mechanisms which includes, enzymes degrading β -lactams, enzymes modifying aminoglycosides, alteration in the binding sites for



Figure 5. Key steps in the genomics-driven antibiotic drug discovery process.

quinolones, array of efflux mechanisms and changes taking place in the outer membrane proteins. Presence of these elements in the *A. baumannii* makes it a well drug-resistant pathogen, due to which designing of an appropriate antimicrobial agent is not easy. In fact, it is very much probable that *A. baumannii* would be resistant to some of the most common first line antibiotics so it is advisable that treatment of the infection should be performed after doing the antimicrobial susceptibility testing. On the other hand, the time lapse while accessing correct treatment may result in the complicated health consequences for a patient's. Presently the carbapenems such as Imipenem are generally given as a drug of preference to a patient's suspected for the serious infections caused by the *Acinetobacter* [63]. Till date, there is incomplete knowledge regarding the clinical response and resistance mechanisms of the antimicrobials [64]. The consistency and comparability of various methods used in the susceptibility testing of *A. baumannii* is again not so truthful [65]. As in the case of β -lactams the tenacity of subtle growth after an obvious end point by broth microdilution is of big concern and this explains its poor reaction with the disc diffusion method. Another novel carbapenem drug Doripenem, is found to active against the susceptible *A. baumannii* [66–68].

While the drug Doripenem was found to be ineffective against isolates of the *A. baumannii* producing blaOXA-23 or blaIMP-4 or metallo- β -lactamases [69].

1.1.9.2. New strategy

Commercially first antibiotic was introduced long back and it did not take long time for microbes to develop resistance for this antibiotics. Now days the extensive use of many antimicrobial drugs offers the ideal conditions for the development and spread of multiple drug resistance in the microbes. Although old days expert researchers successfully identified the ways used by microbes to evade the effect of antibiotics. Presently multidrug resistance is a serious worldwide problem it affects the peoples of all socioeconomic classes. The bacteria can adopt intricate mechanisms to evade the fatal effects of antibiotics and understanding of these mechanisms conferring resistance can guide to design the new drugs. Recognition of the potential for the development of the resistance should garner newfound respect for the urgently innovation of the novel agents crucial for the cure of infectious diseases. Another possible approach is to develop the drugs which will target the enzyme catalyzing the critical steps of metabolic process of bacteria and till date maximum of well known antimicrobials target the most critical metabolic process namely protein synthesis, DNA repair and replication pathway and peptidoglycan (component of the cell wall) biosynthesis pathway. As a result of high natural selection capacity of bacteria the drug resistant strains have been emerged against these members. Hence, new targets from these pathways are needed to be explored for the development of novel and potent classes of antimicrobials. The advancement in the high throughput DNA sequencing resulted in the exponential rise in genome sequence data and this genome sequence data can be utilized for the innovation of new antibacterial scaffolds, which would be a better strategy to keep multidrug-resistant infections at bay. The lack of good therapeutic options is increasingly limited for *Acinetobacter*. Hence the discovery of novel targets from metabolically important pathways may turn to decisive proposal. Here, we discuss about the routine targets and about new targets which are not explored but are crucial for the survival of the organism.

1.1.10. The conventional targets in *Acinetobacter*

The conventional targets of *Acinetobacter* can be divided into three categories first one is the enzymes which inactivate the antibiotics, for example OXA-type inactivating enzymes [70–

72] and some class of metallo- β -lactamases (MBLs). The second one is reduced entry of antibiotics due to efflux pumps or as a result of target site alteration and well known examples of this type targets are porin channels, β -lactamases and efflux pumps [73, 74]. Third category of the conventional targets include alteration of the target or cellular functions due to mutations and most studied examples of this category are *gyrA* and *parC* topoisomerase enzymes. So many research groups are working on the conventional targets of the *Acinetobacter*. In the present study, we have chosen the non-conventional targets, which are essential for the organism survival.

1.1.11. Unconventional target of *Acinetobacter*

Here we have chosen, the targets, Hpa2 and TAG for the structure-based drug designing. The enzyme, Hpa2 is a histone acetyltransferases, (HATs) are known to transfer acetyl group from the acetyl-CoA to the ϵ -amino group of conserved lysine amino acid on histone *N*-terminal tails and core regions or to small molecules such as polyamines antibiotics *etc.* [75]. Generally, acetylation of histone results in the increased DNA accessibility through the neutralization of the positive charge of lysine. This neutralization of the positive charge weakens the interaction between histone and DNA, and in this manner acetylation promotes the transcriptional activation. Furthermore, acetylation is known to be involved in various cellular processes such as DNA repair, DNA replication, bacterial antibiotic resistance, transcriptional silencing and cell cycle progression. Acetylation of histone is also known to regulate and generate binding sites for specific protein-protein interactions [76–80]. Some HATs have the capacity to acetylate a number of growing non-histone substrates. HATs can be grouped into various diversified families based on sequence homology and structural features that show high sequence similarity within families and very less sequence similarity between families. The four major families of the HATs are (i) MYST family including MOZ, Ybf2/Sas3, Sas2, Tip60, (ii) Gcn5 related *N*-acetyltransferase or GNAT family, (iii) p300 family and (iv) Rtt109 family. In spite of very less similarity in primary sequence, most of the HATs possess similar substrate binding motif, and active site structure.

Among all the families of HATs, the members of GNAT (Gcn5-related *N*-acetyl transferases) superfamily illustrates less sequence homology irrespective of remarkably conserved fold of catalytic domain [81]. The acetylation mechanism means how HATs achieve acetyl group transfer, which is described in Figure 6. During the catalysis, acetyl-CoA

and the protein substrate bind with the HAT enzyme to form a ternary complex. First of all, glutamate of the active site supports in the deprotonation of histone lysine, helping in the nucleophilic attack on the carbonyl carbon of the substrate acetyl-CoA. Finally, a tetrahedral intermediate formation takes place, which collapses into the end products *viz.*, CoA and the acetylated protein.

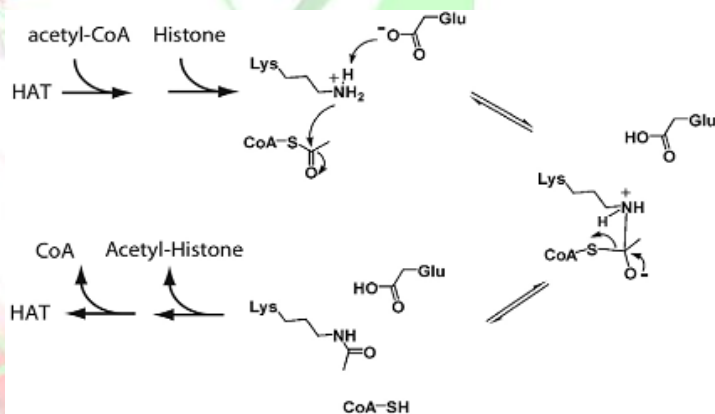


Figure 6. Proposed kinetic and catalytic mechanism of acetyl transfer (Adapted from Ref. [82](#)).

The GNAT family includes Gcn5, PCAF, Hat1, Elp3, Hpa2, Hpa3, ATF-2, and Nut1. Structural analysis performed by Blanchard and co-workers [\[83\]](#) on *S. enteric* showed that aminoglycoside 6'-N-acetyltransferase (AAC6), an enzyme capable of regioselective N-acetylation of antibiotics, exhibits significant primary sequence homology with the GNAT superfamily. This enzyme is found to exist as dimeric unit and its dimer structure is very similar to that of *S. cerevisiae* Hpa2-encoded histone acetyltransferase enzyme [\[84, 85\]](#). The biochemical characterization of enzymes Hpa2 and Hpa3 performed recently [\[86\]](#) illustrates that both of these enzymes have capacity to acetylate histone and some small molecules including the polyamines. Furthermore, they also found that Hpa2 have the capacity to acetylate basic proteins but not the D-amino acids and on contrast to it the protein Hpa3 was found to acetylate D-amino acids [\[87\]](#). To know the possibility of this differential nature of acetylation of different substrates we carried out literature survey to correlate the acetylation of polyamines and function of the Hpa2 enzyme. It is widely reviewed that polyamines are crucial modulators of a variety of ion channels [\[88\]](#) and polyamines are also found to modulate the expression of many genes and protein activity in *Escherichia coli*, which directs

optimal cell growth, viability and defense mechanisms against toxic environmental conditions [83, 86, 89–92].

The addition of polyamines to the growth medium of *P. aeruginosa*, multi-drug resistant bacteria found to enhance the MIC values of many antibiotics (including polycationic antibiotics, aminoglycosides, quinolones, and fluorescent dyes), which indicates the existence of complicated molecular mechanism for polyamine-mediated resistance to multiple antibiotics in *P. aeruginosa* [93, 94]. Additionally, many evidences highlight the significance of polyamines in many key aspects of pathogenesis, survival and virulence of many pathogens [95]. From the above mentioned literature findings, it can be concluded that polyamine acetylation capacity of the Hpa2 might be associated with the antibiotics resistance mechanism. Furthermore, the exact molecular function of the Hpa2 in the prokaryotic as well as in eukaryotic cell is not well studied. As per results showing positive correlation with the hypothesis made during the *in silico* study of Hpa2. We thought to perform the *in vitro* studies on the same enzyme. Therefore, we cloned, expressed and purified the Hpa2 in the *E. coli*. After purification, the 2D and 3D folding of protein were confirmed by the CD spectroscopy, and the oligomeric state was studied using the gel filtration chromatography in presence of various substrates and denaturing agents. Binding constants were calculated using ITC and fluorescence. Crystallization was also attempted to characterize the structure function properties.

Second enzyme taken for the study is TAG from the DNA repair pathway. TAG is DNA glycosylase enzymes which recognize the DNA damage catalyses the first step of base excision repair (BER) pathway (Figure 7) [96].

1.1.11.1. Recognition of DNA Damage

Basically there are three damage-searching mechanisms are proposed for the DNA damage repair enzymes; first one is extrahelical, in this enzyme flips every base into its active site to check for the damage [97–99]. Second one is intrahelical, in which the enzyme spots the intrahelical lesion and then actively flips lesion into its active site; in the third one, the enzyme

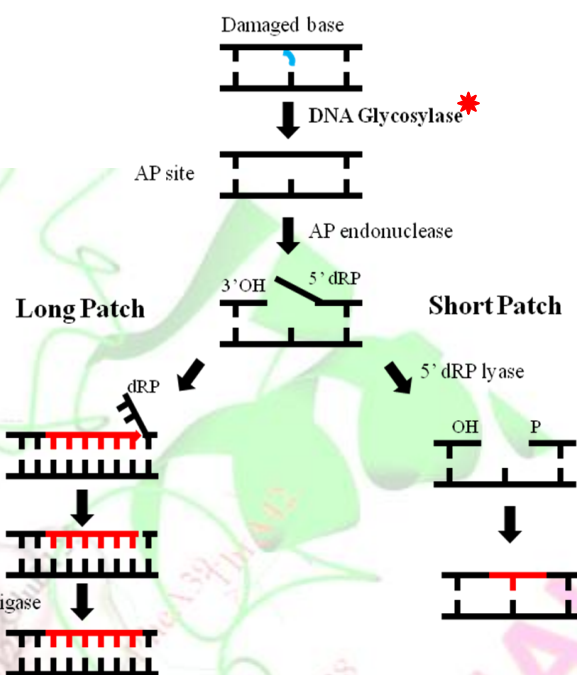


Figure 7. Basic steps of BER pathways.

captures a transiently extrahelical lesion. In the second and third recognition mechanism the enzyme may sense instability of the damaged base pair, as in the case of bulky lesions like ϵ A or thymine distortion in the DNA helical structure is considerable to be noticed. However the perturbations like of mismatched base pairs or methylated lesions cannot be recognized and they need the costly and unfavourable first type of damage repair mechanism for their recognition. The evidences obtained from the high-resolution crystal structures, molecular dynamics and biochemical studies it is known that most of enzymes actively flips base lesions rather than simply capture an extra-helical base.

After recognition the enzyme DNA glycosylases first intercalate the protein residues into the space left by the flipped nucleotide to uphold the base stacking properties. After flipping the damaged bases are recognized, based on size, shape, pattern of hydrogen-bonding, and distribution of electronic charge while fitting into the active site. However normal bases are generally expelled from the enzyme active site due to its steric constraints, and interactions involved in the base recognition and its removal are shown in **Figure 8**.

The enzyme specificities are maintained by the chemical and physical properties of the base-pairing and base binding pocket. After the flipping step, enzyme cleaves the N-glycosylic bond to remove the damaged base.

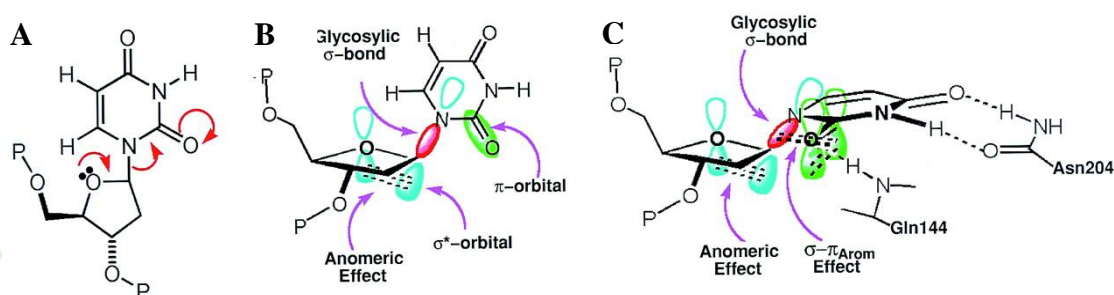


Figure 8. Structure-based reaction mechanism that resolves the apparent orthogonal paradox for electron transpositions by altering the substrate stereochemistry. (A) A simplified valence-bond representation of the glycosidic bond dissociation hides the paradox that the three electron pairs to be transposed are involved in orthogonal orbitals. (B) In the normal *anti*-conformation of deoxyuridine, the σ^* -orbital involved in the anomeric effect and the π -orbital of the C2=O bond are orthogonal to one another, thus preventing orbital overlap. (C) Severe distortions of the deoxyribose and the glycosidic bond in the strained conformation of deoxyuridine enforced by the UDG active center align the pairs of atomic orbitals participating in each electron transposition, thereby electronically coupling the anomeric and σ - π_{Arom} effects to promote bond cleavage. (Adapted from Ref 100).

The Glycosylase enzymes can be classified into two groups; mono-functional and bi-functional enzymes. The mono-functional catalyze a hydrolysis reaction by generating a hydroxyl ion which attacks on the anomeric carbon of the lesioned nucleotide. However, the bi-functional DNA glycosylases enzyme uses an active site amine residue for the nucleophilic attack and proceeds through a Schiff base intermediate. Here we conclude that all glycosylases shares a common mode of action for the recognition of damage than they flips damaged bases out of the DNA helix into its binding pocket, the architecture of which permits chemically and physically detection of base irregularities.

1.1.11.2. 3-Methyladenine DNA glycosylase I (TAG)

In the prokaryotic cells two DNA glycosylases are found they known to cleave the glycosidic bond of alkylated purine bases in DNA via the hydrolysis. They are named as the 3-methyladenine DNA glycosylase II (AlkA), it has been studied in free as well as in the bound form and structural analysis reveals that it belongs to HhH family. It cleaves the modified/mispaired adenine and guanine using water as the activated nucleophile in the hydrolysis. The conservation of structural fold suggests that this scaffold developed in the early stage of the protein evolution scaffold for the DNA binding.

The second alkyl purine DNA glycosylase, 3-MeA DNA glycosylase I (TAG) is the constitutively expressed. Its phylogeny is not well understood due to absence of sequence similarity with other DNA glycosylases, including the trademark HhH motifs of the superfamily. It does not possess sequence as well structural similarity with its structural and functional cousin AlkA.

Furthermore, one interesting thing noticed about the enzyme glycosylase by the Celli and Tsolis [101] in the pathogen is that parasite *H. pylori* colonizes the gastric mucosa are exposed to reactive oxygen and reactive nitrogen species (ROS; RNS). Which are produced by a strong respiratory burst from neutrophils, this exposer generates various types of damaged bases. And as usual the generated single-nucleobase lesions are repaired by the BER pathway. But new interesting thing noticed is that DNA glycosylases do not recognize and excise the mutagenic nucleobase lesions. However, enzyme MutY found to involved in the excision of adenine mispaired with 8-oxoG which are known to having high probability of spontaneous mutations. In this way the various DNA glycosylase enzymes are found to be associated with the adaptation of human pathogens (Figure 9).

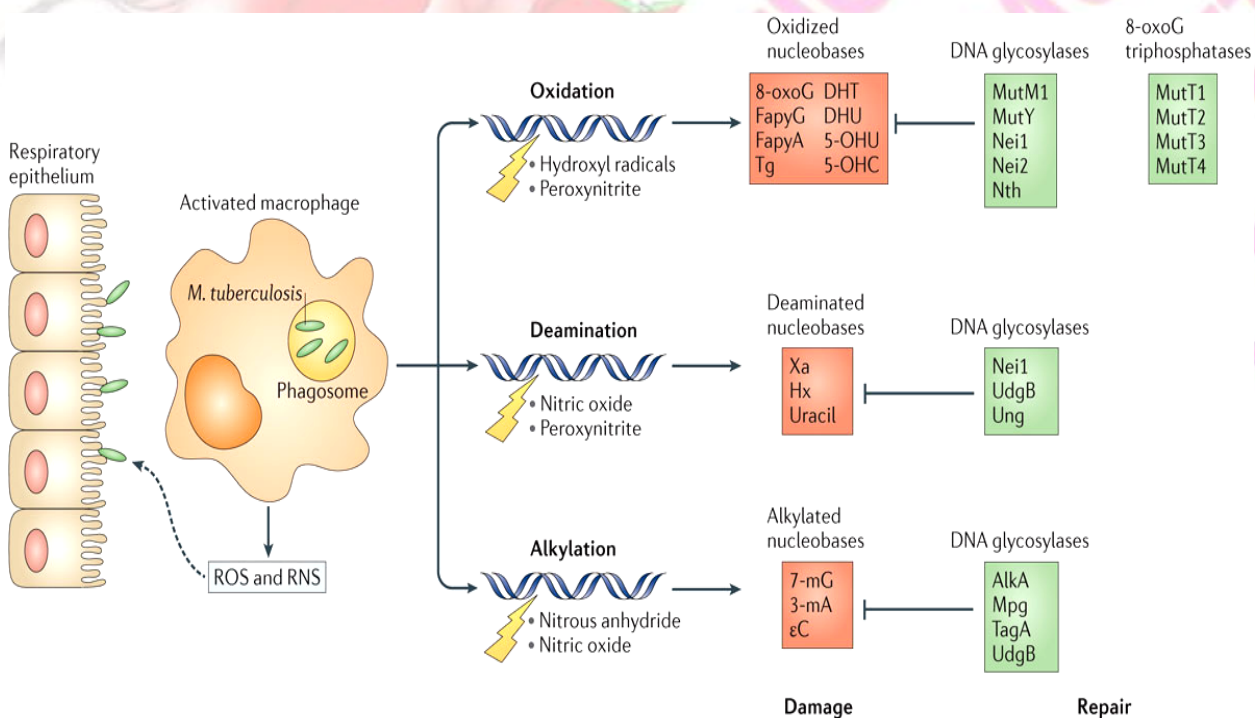


Figure 9. The BER necessities: the repair of DNA damage in human-adapted bacterial pathogens. [101]

The enzyme Hpa2 is known to exist as a dimer in the eukaryotes [102–104]. However, its homologue from prokaryotes is known to exhibit the differential oligomeric states. Recently Minor and his co-workers [105] performed the biochemical studies on the enzyme PA4794 from MDR organism *Pseudomonas aeruginosa*, and found that compounds cefotaxime, cefalotin, cefuroxime, 7-aminocephalosporanic acid, cefmetazole, ceftazidime, cefixime, cephalosporin C, and ammonium 2-(aminocarbonyl)benzoate bind at the active site of enzyme; so these molecules may work as competitive inhibitors [106–113]. Furthermore, PA4794 enzyme was found to acetylate the antibiotics chloramphenicol, which could be related to the antibiotic resistance. Adding to the complexity of possible substrates for the members belonging to GNATs, it appears to be most promiscuous and has capability to acetylate more than one class of substrates [113–118]. Due to all the features described above, especially its absence in the higher organisms and variety in substrate recognition makes it good chemotherapeutic target to be studied.

From the above discussion it is understandable that these enzymes are essential for the *A. baumannii*. The initial steps of this work are *in silico* structure modelling, active site characterization, structural/functional characterization of selected targets and structure-based screening of database to identify novel antibacterial compounds. After that heterologous expression and purification of target proteins are performed to characterize the structure and function of proteins.

1.2. Conclusion

Acinetobacter baumannii is a multiple drug resistant and opportunistic pathogen and is known to cause bacteraemia, secondary meningitis, surgical-site infections, and ventilator associated pneumonia and urinary tract infections. The tactic of improvements in the existing antibiotics is a short term strategy, and has so far been successful at achieving this goal. Consequently, the innovation of new antibacterial scaffolds would be a better strategy to keep multidrug-resistant infections at bay. The lack of good therapeutic options is increasingly limited for *Acinetobacter* so innovation of new antimicrobials is needed to cure the infections caused by the organisms. Our exploration in this direction led us to the search of novel targets which ended at Hpa2 and TAG, in this study both enzymes have been taken to perform the *in silico* and experimental study.

1.3. References

1. Pendleton, J.N., Gorman, S.P., and Gilmore, B.F. (2013) Clinical relevance of the ESKAPE pathogens; *Expert Rev. Anti Infect. Ther.* 3, 297–308.
2. Jones, R.N. (2001) Resistance patterns among nosocomial pathogens: trends over the past few years; *Chest* 119, 397–404.
3. Boucher, H.W., Talbot, G.H., Bradley, J.S., Edwards, Jr, J.E., Gilbert, D., Rice, L.B., Scheld, M., Spellberg, B. and Bartlett, J. (2009) Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America; *Clin. Infect. Dis.* 48, 1–12.
4. Noto, M.J., Boyd, K.L., Burns, W.J., Varga, M.G., Peek, Jr., R.M. and Skaara, E.P. (2015) Toll-Like Receptor 9 Contributes to Defense against *Acinetobacter baumannii* Infection; *Infect. Immun.* 83, 4134–4141.
5. Richards, M.J., Edwards, J.R., Culver, D.H. and Gaynes, R.P. (2000) Nosocomial infections in combined medical-surgical intensive care units in the United States; *Infect Control Hosp. Epidemiol.* 21, 510–515.
6. Livermore, D.M. (1995) beta-Lactamases in laboratory and clinical resistance; *Clin. Microbiol. Rev.* 8, 557–584.
7. Paterson, D.L., Mulazimoglu, L. and Casellas, J.M. *et al.* (2000) Epidemiology of ciprofloxacin resistance and its relationship to extended-spectrum beta-lactamase production in *Klebsiella pneumoniae* isolates causing bacteremia; *Clin. Infect. Dis.* 30, 473–478.
8. Lee, J., Patel, G. and Huprikar, S. *et al.* (2009) Decreased susceptibility to polymyxin B during treatment of carbapenem-resistant *Klebsiella pneumoniae* infection; *J. Clin. Microbiol.* 47, 1611–1612.
9. Bratu, S., Tolaney, P. and Karumudi, U. *et al.* (2005) Carbapenemase-producing *Klebsiella pneumoniae* in Brooklyn, NY: molecular epidemiology and in vitro activity of polymyxin B and other agents; *J. Antimicrob. Chemother.* 56, 128–132.
10. Sakka, S.G., Glauner A.K. and Bulitta J.B. *et al.* (2007) Population pharmacokinetics and pharmacodynamics of continuous versus short-term infusion of imipenem-cilastin in critically ill patients in a randomized, controlled trial; *Antimicrob. Agents Chemother.* 51, 3304–3310.
11. Herrmann, D.J., Peppard, W.J., Ledebor, N.A., Theesfeld, M.L., Weigelt, J.A., and Buechel, B.J. (2008) Linezolid for the treatment of drug-resistant infections; *Expert Rev. Anti Infect. Ther.* 6, 825–48.
12. Livermore, D.M. (2002) Multiple mechanisms of antimicrobial resistance in *Pseudomonas aeruginosa*: our worst nightmare?; *Clin. Infect. Dis.* 34, 634–640.
13. Falagas, M.E. and Karageorgopoulos, D.E. (2008) Pandrug resistance (PDR), extensive drug resistance (XDR), and multidrug resistance (MDR) among Gram-negative *bacilli*: need for international harmonization in terminology; *Clin. Infect. Dis.* 46, 1121–1122.
14. Simor, A.E., Lee, M., Vearncombe, M., JonesPaul, L., Barry, C. and Gomez, M. *et al.* (2002) An outbreak due to multiresistant *Acinetobacter baumannii* in a burn unit: risk factors for acquisition and management; *Infect. Control Hosp. Epidemiol.* 23, 261–267.

15. Gusten, W.M., Hansen, E.A. and Cunha, B.A. (2002) *Acinetobacter baumannii* pseudomeningitis; Heart Lung 31, 76–78.
16. Montefour, K., Frieden, J., Hurst, S., Helmich, C., Headley, D. and Martin, M. (2008) *Acinetobacter baumannii*: an emerging multidrug-resistant pathogen in critical care; Crit. Care Nurse 28, 15–25.
17. Falagas, M.E. and Kopterides, P. (2006) Risk factors for the isolation of multidrug resistant *Acinetobacter baumannii* and *Pseudomonas aeruginosa*: a systematic review of the literature; J. Hosp. Infect. 64, 7–15.
18. Landman, D., Quale, J. M., Mayorga, D., Adedeji, A., Vangala, K. and Ravishankar, J. et al. (2002) Citywide clonal outbreak of multiresistant *Acinetobacter baumannii* and *Pseudomonas aeruginosa* in Brooklyn, NY: the preantibiotic era has returned; Arch. Intern. Med. 162, 1515–1520.
19. Cisneros, J.M., RodríguezBaño, J., FernándezCuenca, F., Ribera, A., Vila, J., Pascual, A. and MartínezMartínez, L. et al. (2005) Risk factors for the acquisition of imipenem resistant *Acinetobacter baumannii* in Spain a nationwide study; Clin. Microbiol. Infect. 11, 874–879.
20. Medina, J., Formento, C., Pontet, J., Curbelo, A., Bazet, C. and Gerez, J. et al. (2007) Prospective study of risk factors for ventilator associated pneumonia caused by *Acinetobacter* species; J. Crit. Care 22, 18–27.
21. Katsaragakis, S., Markogiannakis, H., Toutouzas, K. G., Drimousis, P., Larentzakis, A., Theodoraki, E. M., et al. (2008) *Acinetobacter baumannii* infections in a surgical intensive World; J. Surg. 32, 1194–1202.
22. Bergogne-Berezin, E. and Towner, K.J. (1996) *Acinetobacter* spp. as nosocomial pathogens: microbiological, clinical, and epidemiological features; Clin. Microbiol. Rev. 9, 148–165.
23. Bouvet, P.J.M. and Grimont. P.A.D (1986) Taxonomy of the genus *Acinetobacter* with therecognition of *Acinetobacter baumannii* sp. nov., *Acinetobacter haemolyticus* sp. nov., *Acinetobacter johnsonii* sp. nov., *Acinetobacter junii* sp. nov. & emended description of *Acinetobacter calcoaceticus* and *Acinetobacter lwoffii*; Int. J. Syst. Bacteriol. 36, 228–240.
24. Kuo, L., Teng, L., Yu, C., Ho, S. and Hsueh, P. (2004) Dissemination of a clone of unusual phenotype of pandrug-resistant *Acinetobacter baumannii* at a University Hospital in Taiwan; J. Clin. Microbiol. 42, 1759–1763.
25. Go, E.S., Urban, C., Burns, J., Kreiswirth, B., Eisner, W. and Mariano, N.F. (1994) Clinical and molecular epidemiology of *Acinetobacter* infections sensitive only to polymyxin B and sulbactam; Lancet 12, 1329–1332.
26. Hsueh, P.R., Teng, L.J., Chen, C.Y., Chen, W.H., Yu, C.J., Ho, S.W. and Luh, K.T. (2002) Pandrug-Resistant *Acinetobacter baumannii* causing nosocomial infections in a University in Taiwan; Emerg. Infect. Dis. 8, 827–832.
27. Euzéby, J.P. (2008) List of prokaryotic names with standing in nomenclature <http://www.bacterio.cict.fr/classificationac.html#Acinetobacter>.

28. Jawad, A., Heritage, J., Snelling, A.M., Gascoyne-Binzi, D.M. and Hawkey, P.M. (1996) Influence of relative humidity and suspending menstrual on survival of *Acinetobacter* spp on dry surfaces; *J. Clin. Microbiol.* 34, 2881–2887.
29. Getchell-White, S.I., Donowitz, L.G. and Groeschel, D.H.M. (1989) the inanimate environment of an intensive care unit as a potential source of nosocomial bacteria: evidence for long survival of *Acinetobacter calcoaceticus*; *Infect. Control Hosp. Epidemiol.* 10, 402–407.
30. Kramer, A., Schwebke, I. and Kampf, G. (2006) How long do nosocomial pathogens persist on inanimate surfaces? A systematic review; *BMC Infect. Dis.* 6, 130.
31. Cappelli, G., Sereni, L., Scialoja, M.G., Morselli, M., Perrone, S., Ciuffreda, A., Bellesia, M., Inguaggiato, P., Albertazzi, A. and Tetta, C. (2003) Effects of biofilm formation on haemodialysis monitor disinfection; *Nephrol. Dial. Transplant.* 18, 2105–2111.
32. Loukili, N.H., Granbastien, B., Faure, K., Guery, B. and Beaucaire, G. (2006) Effect of different stabilized preparations of peracetic acid on biofilm; *J. Hosp. Infect.* 63, 70–72.
33. Cuhna, B.A. *Acinetobacter* (2007) <http://www.emedicine.com/MED/topic3456.htm>
34. Allen, K.D. and Green, H.T. (1987) Hospital outbreak of multi-resistant *Acinetobacter anitratus*: an airborne mode of spread?; *J. Hosp. Infect.* 9, 110–119.
35. Falagas, M.E., Bliziotis, I.A. and Siempos, I.I. (2006) Attributable mortality of *Acinetobacter baumannii* infections in critically ill patients: a systematic review of matched cohort and case-control studies; *Crit. Care* 10, R48.
36. Wang, J.T., McDonald, L.C., Chang, S.C. and Ho, M. (2002) Community-Acquired *Acinetobacter baumannii* bacteremia in adult patients in Taiwan; *J. Clin. Microbiol.* 40, 1526–1529.
37. Perez, F., Hujer, A.M., Hujer, K.M., Decker, B.K., Rather, P.N. and Bonomo, R.A. (2007) Minireview: Global challenge of multidrug-resistant *Acinetobacter baumannii*; *Antimicrob. Agents and Chemother.* 51, 3471–3484.
38. Brown, S., Bantar, C., Young, H. and Amyes, S. (1996) An outbreak of imipenem resistance in *Acinetobacter* strains from Buenos Aires, Argentina, abstract C-122, p 56 In Abstracts of the 36th Interscience Conference on Antimicrobial Agents and Chemotherapy; American Society for Microbiology, Washington, DC.
39. Clatworthy, A.E., Pierson, E. and Hung, D. T. (2007) Targeting virulence: a new paradigm for antimicrobial therapy; *Nat. Chem. Biol.* 3, 541–548.
40. Fernandes, P. (2006) Antibacterial discovery and development-the failure of success?; *Nat. Biotechnol.* 24, 1497–1503.
41. Overbye, K.M. and Barrett, J.F. (2005) Antibiotics: where did we go wrong?; *Drug Discov. Today* 10, 45–52.
42. Benton, B., Breukink, E., Visscher, I., Debabov, D., Lunde, C., Janc, J., Mammen, M. and Humphrey, P. (2007) Telavancin inhibits peptidoglycan biosynthesis through preferential targeting of transglycosylation: evidence for a multivalent interaction between telavancin and lipid II; *Int. J. Antimicrob. Agents* 29, 51–52.

43. Leach, K.L., Swaney, S.M., Colca, J.R., McDonald, W.G., Blinn, J.R., Thomasco, L.M., Gadwood, R.C., Shinabarger, D., Xiong, L. and Mankin, A.S. (2005) The site of action of oxazolidinone antibiotics in living bacteria and in human mitochondria; *Mol. Cell* 26, 393–402.
44. Strohl, W.R. (1997) *Biotech Antibiotics*. Marcel Dekker Inc., New York, USA.
45. Straus, S.K. and Hancock, R.E.W. (2006) Mode of action of the new antibiotic for Gram-positive pathogens daptomycin: comparison with cationic antimicrobial peptides and lipopeptide; *Biochim. Biophys. Acta* 1758, 1215–1223.
46. Fisher, J.F. and Mobashery, S. (2010) *Enzymology of Bacterial Resistance*. *Comprehensive Natural Products II*. In: *Enzymes and Enzyme Mechanisms*, vol. 8. Elsevier, pp. 443–487.
47. Yang, W., Moore, I.F., Koteva, K.P., Bareich, D.C., Hughes, D.W. and Wright, G.D. (2004) TetX is a flavin-dependent monooxygenase conferring resistance to tetracycline antibiotics; *J. Biol. Chem.* 279, 52346–52352.
48. Bonnet, R. (2004) Growing group of extended-spectrum beta-lactamases: the CTX-M enzymes; *Antimicrob. Agents Chemother.* 48, 1–14.
49. Blanchard, A. (2004) Bacterial acetyltransferase capable of regioselective N-acetylation of antibiotics and histones; *Chem. Biol.* 11, 565–573.
50. Schwarz, S., Kehrenberg, C., Doublet, B. and Cloeckaert, A. (2004) Molecular basis of bacterial resistance to chloramphenicol and florfenicol; *FEMS Microbiol. Rev.* 28, 519–542.
51. Matsuoka, M. and Sasaki, T. (2004) Inactivation of macrolides by producers and pathogens; *Curr. Drug Targets Infect. Disord.* 4, 217–240.
52. Brisson-Noel, A., Delrieu, P., Samain, D. and Courvalin, P. (1988) Inactivation of lincosaminide antibiotics in *Staphylococcus*: identification of lincosaminide O-nucleotidyltransferases and comparison of the corresponding resistance genes; *J. Biol. Chem.* 263, 15880–15887.
53. Spratt, B.G. (1994) Resistance to antibiotics mediated by target alterations; *Science* 264, 388–393.
54. Ruiz, J. (2003) Mechanisms of resistance to quinolones: target alterations, decreased accumulation and DNA gyrase protection; *J. Antimicrob. Chemother.* 51, 1109–1117.
55. Wolter, D.J., Hanson, N.D. and Lister, P.D. (2004) Insertional inactivation of *oprD* in clinical isolates of *Pseudomonas aeruginosa* leading to carbapenem resistance; *FEMS Microbiol. Lett.* 236, 137–143.
56. Bush, K. and Fisher, J.F. (2011) Epidemiological expansion, structural studies, and clinical challenges of new β -lactamases from Gram-negative bacteria; *Annu. Rev. Microbiol.* 65, 455–478.
57. Woodford, N., Turton, J.F. and Livermore, D.M. (2011) Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance; *FEMS Microbiol. Rev.* 35, 736–755.
58. Albertson, D.G. (2006) Gene amplification in cancer; *Trends Genet.* 22, 447–455.

59. Nicoloff, H., Perreten, V., McMurry, L.M. and Levy, S.B. (2006) Role for tandem duplication and lon protease in AcrAB-TolC-dependent multiple antibiotic resistance (Mar) in an *Escherichia coli* mutant without mutations in *marRAB* or *acrRAB*; *J. Bacteriol.* 188, 4413–4423.
60. Nicoloff, H., Perreten, V. and Levy, S.B. (2007) Increased genome instability in *Escherichia coli* ion mutants: relation to emergence of multiple antibiotic resistant (Mar) mutants caused by insertion sequence elements and large tandem genomic amplifications; *Antimicrob. Agents Chemother.* 51, 1293–1303.
61. Matthews, P.R. and Stewart, P.R. (1988) Amplification of a section of chromosomal DNA in methicillin-resistant *Staphylococcus aureus* following growth in high concentrations of methicillin; *J. Gen. Microbiol.* 134, 1455–1464.
62. Damien, M. and Rosenberg, M. (2001) Exploiting Genomics to Discover New Antibiotics; *TRENDS Microbiol.* 9, 611–617.
63. Cisneros, J.M. and Rodriguez-Bano, J. (2002) Nosocomial bacteremia due to *Acinetobacter baumannii*: epidemiology, clinical features and treatment; *Clin. Microbiol. Infect.* 8, 687–693.
64. Kahlmeter, G., Brown, D.F., Goldstein, F.W., MacGowan, A.P., Mouton, J.W., Odenholt, I., Rodloff, A., Soussy, C.J., Steinbakk, M., Soriano, F. and Stetsiouk, O. (2006) European Committee on Antimicrobial Susceptibility Testing (EUCAST) technical notes on antimicrobial susceptibility testing; *Clin. Microbiol. Infect.* 12, 501–503.
65. Swenson, J.M., Killgore, G.E. and Tenover, F.C. (2004) Antimicrobial susceptibility testing of *Acinetobacter* spp. by NCCLS broth microdilution and disk diffusion methods; *J. Clin. Microbiol.* 42, 5102–5108.
66. Jones, R.N., Huynh, H.K., Biedenbach, D.J., Fritsche, T.R. and Sader, H.S. (2004) Doripenem (S-4661), a novel carbapenem: comparative activity against contemporary pathogens including bactericidal action and preliminary in vitro methods evaluations; *J. Antimicrob. Chemother.* 54, 144–154.
67. Jones, R.N., Sader, H.S. and Fritsche, T.R. (2005) Comparative activity of doripenem and three other carbapenems tested against Gram-negative bacilli with various beta-lactamase resistance mechanisms; *Diagn. Microbiol. Infect. Dis.* 52, 71–74.
68. Fritsche, T.R., Stilwell, M.G. and Jones, R.N. (2005) Antimicrobial activity of doripenem (S-4661): a global surveillance report; *Clin. Microbiol. Infect.* 11, 974–984.
69. Mushtaq, S., Ge, Y. and Livermore, D.M. (2004) Comparative activities of doripenem versus isolates, mutants, and transconjugants of Enterobacteriaceae and *Acinetobacter* spp. with characterized β -lactamases; *Antimicrob. Agents Chemother.* 48, 1313–1319.
70. Brown, S. and Amyes, S. (2006) OXA β -lactamases in *Acinetobacter*: the story so far; *J. Antimicrob. Chemother.* 57, 1–3.
71. Kaur, N., Khokhar, M., Jain, V., Bharatam, P.V., Sandhir, R. and Tewari, R. (2013) Identification of druggable targets for *Acinetobacter baumannii* via subtractive genomics and plausible Inhibitors for MurA and MurB; *Appl. Biochem. Biotech.* 171, 417–436.

72. Adane, L., Bharatam, P.V., Sharma, V. (2010) A common feature-based 3D-pharmacophore model generation and virtual screening: identification of potential PfDHFR inhibitors; *J. Enzy. Inhibition Med. Chem.* 25, 635–645.
73. Thomson, J.M. and Bonomo, R.A. (2005) The threat of antibiotic resistance in Gram-negative pathogenic bacteria: beta-lactams in peril; *Curr. Opin. Microbiol.* 8, 518–24.
74. Mussi, M.A., Limansky, A.S. and Viale, A.M. (2005) Acquisition of resistance to carbapenems in multidrug-resistant clinical strains of *Acinetobacter baumannii*: natural insertional inactivation of a gene encoding a member of a novel family of beta-barrel outer membrane proteins; *Antimicrob. Agents Chemother.* 49, 1432–440.
75. Li, J., Nation, R.L., Milne, R.W., Turnidge, J.D. and Coulthard, K. (2005) Evaluation of colistin as an agent against multi-resistant Gram-negative bacteria; *Int. J. Antimicrob. Agents* 25, 11–25.
76. Tsiang, M., Jones, G., Niedziela-Majka, A., Hung, M., Jin, D., Hu, E., Yant, S., Samuel, D., Liu, X. and Sakowicz, R. (2012) A trimer of dimers is the basic building block for human immunodeficiency virus-1 capsid assembly; *Biochemistry* 51, 4416–4428.
77. Chandra, V., Jasti, J., Kaur, P., Srinivasan, A., Betzel, C. and Singh, T.P. (2002) Structural basis of phospholipase A2 inhibition for the synthesis of prostaglandins by the plant alkaloid aristolochic acid from a 1.7 Å crystal structure; *Biochemistry* 41, 10914–10919.
78. Mohammadi, M., Chalavi, V., Novakova-Sura, M., Laliberte, J.F. and Sylvestre, M. (2007) Expression of bacterial biphenyl- α -chlorobiphenyl dioxygenase genes in tobacco plants; *Biotech. Bioeng.* 97, 496–505.
79. Dai, S., Vaillancourt, F.H., Maaroufi, H., Drouin, N.M., Neau, D.B., Snieckus, V., Bolin, J.T. and Eltis, L.D. (2002) Identification and analysis of a bottleneck in PCB biodegradation; *Nat. Struct. Mol. Biol.* 9, 934–939.
80. Tarafdar, P.K., Vedantam, L.V., Sankhala, R.S., Purushotham, P., Podile, A.R. and Swamy, M.J. (2014) Oligomerization, conformational stability and thermal unfolding of harpin, HrpZPss and its hypersensitive response-inducing C-terminal fragment, C-214-HrpZPss; *PLoS ONE* 9, e109871/1–18.
81. Loidl, P. (1994) Histone acetylation: Facts and questions; *Chromosoma* 103, 441–449.
82. Berndsen, C.E. and Denu, J.M. (2008) Catalysis and substrate selection by histone/protein lysine acetyltransferases; *Curr. Opin. Struct. Boil.* 18, 682–689.
83. Vetting, M.W., Magnet, S., Nieves, E., Roderick, S.L. and Blanchard, J.S. (2004) A bacterial acetyltransferase capable of regioselective *N*-acetylation of antibiotics and histones; *Chem. Biol.* 11, 565–573.
84. Kuo, M.H. and Allis, C.D. (1998) Roles of histone acetyltransferases and deacetylases in gene regulation; *Bioessays*. 20, 615–626.
85. Betzel, C., Gourinath, S., Kumar, P., Kaur, P., Perbandt, M., Eschenburg, S. and Singh, T.P. (2001) Structure of a serine protease proteinase K from *Triticachium album limber* at 0.98 Å resolution; *Biochemistry* 40, 3080–3088.

86. Sampath, V., Liu, B., Tafrov, S., Srinivasan, M., Rieger, R., Chen, E.I. and Sternglanz, R. (2013) Biochemical characterization of Hpa2 and Hpa3-two small closely related acetyltransferases from *S. cerevisiae*; *J. Biol. Chem.* 288, 21506–21513.
87. Hong, L., Schroth, G.P., Matthews, H.R., Yau, P. and Bradbury, E.M. (1993) Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA; *J. Biol. Chem.* 268, 305–314.
88. Norton, V.G., Imai, B.S., Yau, P. and Bradbury, E.M. (1989) Histone acetylation reduces nucleosome core particle linking number change; *Cell* 57, 449–457.
89. Strahl, B.D. and Allis, C.D. (2000) the language of covalent histone modifications; *Nature* 403, 41–45.
90. Williams, K. (1997) Modulation and block of ion channels: A new biology of polyamines; *Cell. Signal.* 9, 1–13.
91. Pan, Y.H., Liao, C.C., Kuo, C.C., Duan, K.J., Liang, P.H., Yuan, H.S., Hu, S.T. and Chak, K.F. (2006) The critical roles of polyamines in regulating ColE7 production and restricting ColE7 uptake of the colicin-producing *Escherichia coli*. *J. Biol. Chem.* 281, 13083–13091.
92. Chattopadhyay, M.K., Tabor, C.W. and Tabor, H. (2003) Polyamines protect *Escherichia coli* cells from the toxic effect of oxygen; *Proc. Natl. Acad. Sci. USA* 100, 2261–2265.
93. Rato, C., Amirova, S.R., Bates, D.G., Stansfield, I. and Wallace, H.M. (2011) Translational recoding as a feedback controller: Systems approaches reveal polyamine-specific effects on the antizyme ribosomal frameshift; *Nucleic Acids Res.* 39, 4587–4597.
94. Kwon, D.H. and Lu, C-D. (2006) Polyamines induce resistance to cationic peptide, aminoglycoside, and quinolone antibiotics in *Pseudomonas aeruginosa* PAO1; *Antimicrob. Agents Chemother.* 50, 1615–1622.
95. Shah, P. and Swiatlo, E. (2008) A multifaceted role for polyamines in bacterial pathogens; *Mol. Microbiol.* 68, 4–16.
96. Nickson, C.M. and Parsons, J.L. (2014) Monitoring regulation of DNA repair activities of cultured cells in-gel using the comet assay; *Front. Genet.* 5, 232–241.
97. Peralta-Videoa, J.R., Lopez, M.L., Narayan, M., Saupe, G., Gardea-Torresdey, J. (2009) The biochemistry of environmental heavy metal uptake by plants: implications for the food chain; *Int. J. Biochem. Cell Biol.* 41, 1665–1677.
98. Choudhury, C., Priyakumar, U.D., Sastry, G.N. (2015) Dynamics Based Pharmacophore Models for Screening Potential Inhibitors of Mycobacterial Cyclopropane Synthase; *J. Chem. Inf. Model.* 55, 848–860.
99. Venkatesh, R., Ramaiah, M.J., Gaikwad, H.K., Janardhan, S., Bantu, R., Nagarapu, L., Sastry, G.N. Ganesh, A.R. and Bhadra, M. (2015) Luotonin-A based quinazolinones cause apoptosis and senescence via HDAC inhibition and activation of tumor suppressor proteins in HeLa cells; *Eur. J. Med. Chem.* 94, 87–101.
100. Sudip S. Parikh, S.S, Walcher, G, Jones, G.D., Slupphaug, G., Krokan, H.E., Blackburn, G.M. and Tainer, J.A (2000) Uracil-DNA glycosylase–DNA substrate and product

- structures: Conformational strain promotes catalytic efficiency by coupled stereoelectronic effects Proc. Natl. Acad. Sci. USA 97, 5083–5088.
101. Celli, J. and Tsolis, R.M. (2015) Bacteria, the endoplasmic reticulum and the unfolded protein response: friends or foes?; Nat. Rev. Microbiol. 13, 83–94.
 102. Tsiang, M., Jones, G.S., Hung, M., Samuel, D., Novikov, N., Mukund, S., Brendza, K.M., Niedziela-Majka, A., Jin, D., Liu, X., Mitchell, M., Sakowicz, R. and Geleziunas, R. (2011) Dithiothreitol causes HIV-1 integrase dimer dissociation while agents interacting with the integrase dimer interface promote dimer formation; Biochemistry 50, 1567–1581.
 103. Betzel, C., Pal, G.P. and Saenger, W. (1988) Three-dimensional structure of proteinase K at 0.15-nm resolution; Eur. J. Biochem. 178, 155–171.
 104. Sylvestre, M., Macek, T. and Mackova, M. (2009) Transgenic plants to improve rhizoremediation of polychlorinated biphenyls (PCBs); Current Opinion Biotech. 20, 242–247.
 105. Majorek, K.A., Kuhn, M.L., Chruszcz, M., Anderson, W.F. and Minor, W. (2013) Structural, functional and inhibition studies of a GNAT superfamily protein PA4794: a new C-terminal lysine protein acetyltransferase from *Pseudomonas aeruginosa*; J. Biol. Chem. 42, 30223–30235.
 106. Ahmad, E., Naeem, A., Javed, S., Yadav, S. and Khan, R.H. (2007) The minimal structural requirement of concanavalin A that retains its functional aspects; J. Biol. Chem. 142, 307–315.
 107. Mishra, V., Bilgrami, S., Sharma, R.S., Kaur, P., Yadav, S., Krauspenhaar, R., Betzel, C., Voelter, W., Babu, C.R. and Singh, T.P. (2005) Crystal structure of himalayan mistletoe ribosome-inactivating protein reveals the presence of a natural inhibitor and a new functionally active sugar-binding site; J. Biol. Chem. 280, 20712–20721.
 108. Yadav, S., Tomar, A.K., Jithesh, O., Khan, M.A., Yadav, R.N., Srinivasan, A., Singh, T. P. and Yadav, S. (2011) Purification and Partial Characterization of Low Molecular Weight Vicilin-Like Glycoprotein from the Seeds of *Citrullus lanatus*; Protein J. 30, 575–580.
 109. Dhaked, R.K., Alam, S.I. and Singh, L. (2005) Characterization of beta-galactosidase from an Antarctic *Bacillus* sp; Indian J. Biotech. 4, 227–231.
 110. Mishra, V., Ethayathulla, A.S., Sharma, R.S., Yadav, S., Krauspenhaar, R., Betzel, C., Babu, C.R. and Singh, T.P. (2004) Structure of a novel ribosome-inactivating protein from a hemi-parasitic plant inhabiting the northwestern Himalayas; Acta Crystallogr. D: Biol. Crystallogr. 60, 2295–2304.
 111. Mishra, V., Bilgrami, S., Sharma, R.S., Kaur, P., Yadav, S., Krauspenhaar, R., Betzel, C., Voelter, W., Babu, C.R. and Singh, T.P. (2005) Crystal structure of himalayan mistletoe ribosome-inactivating protein reveals the presence of a natural inhibitor and a new functionally active sugar-binding site; J. Biol. Chem. 280, 20712–20721.
 112. Deffuant, G., Neau, D., Amblard, F. and Weisbuch, G. (2000) Mixing beliefs among interacting agents; Adv. Complex Syst. 3, 87–98.

113. Narahari, A., Singla, H., Nareddy, P.K.; Bulusu, G., Surolia, A. and Swamy, M.J. (2011) Isothermal Titration Calorimetric and Computational Studies on the Binding of Chitooligosaccharides to Pumpkin (*Cucurbita maxima*) Phloem Exudate Lectin; *J. Phys. Chem. B* 115, 4110–4117.
114. Kuhn, M.L., Majorek, K.A., Minor, W. and Anderson, W.F. (2013) Broad-substratescreen as a tool to identify substrates for bacterial Gcn5-related N-acetyltransferases with unknown substrate specificity; *Protein Sci.* 22, 222–230.
115. Vetting, M.W., Magnet, S., Nieves, E., Roderick, S. L. and Blanchard, J.S. (2004) A bacterial acetyltransferase capable of regioselective N-acetylation of antibiotics and histones; *Chem. Biol.* 11, 565–573.
116. Kim, C., Villegas-Estrada, A., Heseck, D. and Mobashery, S. (2007) Mechanistic characterization of the bifunctional aminoglycoside-modifying enzyme AAC (3)-Ib/AAC(6')-Ib' from *Pseudomonas aeruginosa*; *Biochemistry* 46, 5270–5282.
117. Sylvestre, M., Mailhiot, K., Ahmad, D. and Massé, R. (1989) Isolation and preliminary characterization of a 2-chlorobenzoate degrading *Pseudomonas*; *Can. J. Microbiol.* 35, 439–443.
118. Adane, L., Patel, D.S. and Bharatam, P.V. (2010) Shape- and chemical feature-based 3D-pharmacophore model generation and virtual screening: Identification of Potential Leads for *P. falciparum* DHFR Enzyme Inhibition; *Chem. Biol. Drug Design* 75, 115–126.

2.1. Introduction

Knowledge of protein three-dimensional structure is crucial for structural biologist. The 3D structure of a protein provides valuable information including functional annotation, evolutionary features and structural information needed in the rational drug designing. Experimental techniques generally used to characterize protein structures such as NMR spectroscopy or X-ray crystallography are difficult, time consuming and not successful with all the proteins, causing an ever increasing gap between the numbers of protein sequences available and solved protein structures (Figure 1). Though the rate of structure determination through the experimental methods will continue to increase, the number of newly discovered sequences grows much faster than the number of structures solved.

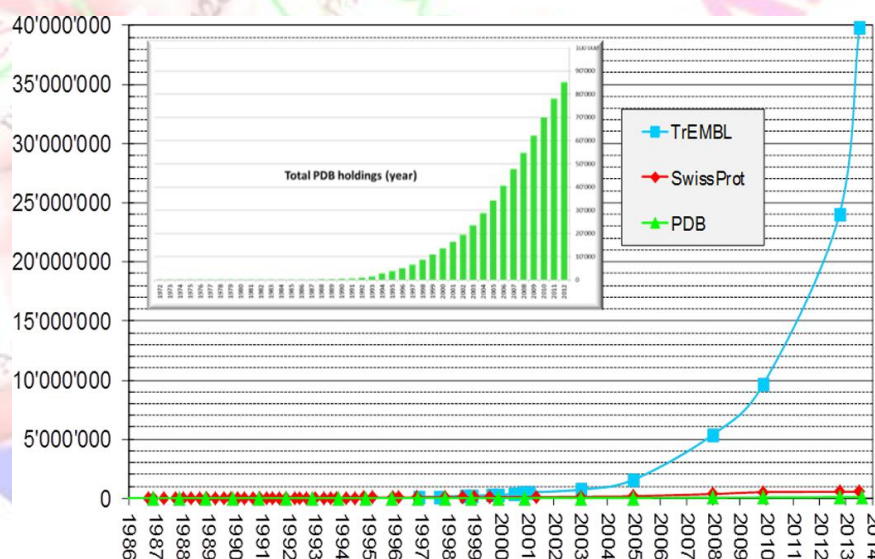


Figure 1. Gap between the sequences and the structures of proteins. The number of entries in the SwissProt and trEMBL sequence databases (UniProtConsortium, 2013) and the PDB [1] are growing exponentially, while the “protein structure gap” between sequence and structures is widening dramatically. Inset: Growth of PDB holdings from 1972 to 2013.

Approximately 11,0350 protein structures are available in the Protein Data Bank (PDB), <http://www.rcsb.org/pdb> (2015), while ~ 40,000,000 sequences are deposited in the SwissProt and trEMBL sequence databases. The computational protein structure prediction methods are widely used to fill the large gap present between the numbers of protein

sequences to its 3D structures and to achieve protein three-dimensional structures with resolution comparable to experimental results. Due to efforts made by the experimental structural biology community a rising fraction of protein families has at least one member in the PDB database solved experimentally [1]. Furthermore, development of accurate profile HMM methods permit the successful detection of remote templates [2]. As a result, during the last decade, a paradigm shift has occurred and nowadays structural information, for at least one member of each family, either from experimental or computational sources is available. Generally, computational models represent only fractions of a protein and one of the unanswered questions in comparative modeling is how to merge the information obtained from multiple templates. Development of reproducible and trustworthy refinement methods having consistently improvement ability by shifting the coordinates closer to the native state are needed to enhance the quality of models.

2.1.1. Protein structure prediction by comparative homology modeling

Pure computational methods (*de novo*), aimed to build the protein native 3D structure from its sequence of amino acids directly without any experimental data, appear to be highly insufficient to provide high-resolution 3D structural for the majority of proteins [3, 4]. To overcome these problems, integrative structure solution techniques, which can combine experimental data having conflicting or ambiguous information from various sources, are indispensable. Particularly, the computational modeling combined with experimental information has proven powerful [5, 6]. However, as shown in Figure 1, the sensitive and sophisticated next-generation sequencing techniques are available causing widening in the protein structure gap. Fortunately, proteins having evident sequence similarity comprise similar three-dimensional structures. In homologous proteins, structural diversity is found to increase with evolutionary distance as noticed and presented by Chothia and Lesk [7]. Based on this fact, methodology for comparative modeling of protein structures was developed two decades ago, which extrapolates available 3D protein structure information derived experimentally, to so far uncharacterized protein sequences [8–11]. In present scenario, the comparative homology modeling techniques have settled into fully automated stable pipelines which are successful in generating trustworthy and tremendous number of 3D models. Since comparative homology modeling methods use the fact that evolutionary related proteins share a similar structure hence models of a protein with unknown structure can be built based on an

alignment of a protein of known structure. Today, homology modeling techniques are emerged as a fully automated stable pipeline which generates reliable three-dimensional models.

2.1.2. Theoretical basis of comparative homology modeling

The protein structure is found to be more conserved than sequence and similarity in sequence suggests similarity in the structure, facilitating the prediction of proteins structure with similar sequences. During the course of evolution acquired mutations are constrained to preserve protein intermolecular and intramolecular interactions responsible for the designated functions in protein families and superfamilies [12]. The study performed on the 25 proteins from eight families found the presence of highly conserved structural regions [7]. The homologous proteins having sequence identity greater than 50% showed more than 90% structurally superimposed backbone C α atoms, and when the sequence identity is decreased to ~20%, only 42% backbone C α atoms showed structural similarity. The obtained rmsd deviation for >50% and ~20% sequence conservation segments of protein measured to be ~1 Å and ~3 Å, respectively. This quantitative analysis states that success rate in prediction of a correct target protein structure from a template protein structure depends on the sequence similarity. Generally, comparative modeling is reserved to the confident homology detection region. Empirical methods based on similarity of amino acid sequence with template have shown good performance, and for this reason, they are also known as template-based approaches. These template-based methods are especially accurate and be able to predict structures similar to the native structure. However, application of template-based approaches is limited to the target protein for which a homologous protein with structure solution is available in the database. Application of this method is further extended by a better coverage of protein structural space and by increase in the computational power. In a recent study, Kryshtafovych and co-workers noticed a positive correlation between prediction accuracy of comparative homology modeling method and sequence similarity among target and template protein [13]. The minimum percentage of sequence similarity needed to deduce structural similarity has been well characterized [14]. Now it is widely accepted that if target and template have sequence identity more than 35% for alignments of ~100 residues then template-based structure prediction methods are suitable to use. Accurate models are appropriate for a wide range of biomolecular applications such as protein active sites prediction, computation of

effect of mutations on the protein structure and function, and structure-based virtual screening. Particularly, comparative modeling has enabled structure-based drug design against protein targets with unknown structures.

2.1.3. Steps in comparative modeling

Conventional homology modeling typically involves four steps: (1) identification of homologous protein structure that can be used as template(s) for modeling, (2) pairwise alignment of the target sequence to the template(s), (3) building a model for the target sequence based on the information from the alignment(s), and (4) evaluation of the model. These all steps can be repeated until a suitable model is obtained. Homology modeling starts with fold assignment and template identification and if the target sequence has high sequence identity (>35%) with the structure then template finding is quite straight forward and often it is completed with the dynamic programming methods. BLAST is simplest searching method, which is based on a pairwise sequence comparison of target and template. Problems associated with the homologue detection and target-template sequence alignment are described in the following section.

2.1.4. Limitations due to low sequence identity

When sequence identity is higher than 40%, pairwise sequence alignment is easy as there are not many gaps; about 90% of main-chain atoms could be modelled with an rmsd error of about 1 Å [15]. In this range of sequence identity, difference in structure between target and template proteins mostly arises from loops and side-chains. When sequence identity lies between 30–40%, getting correct alignment of sequence becomes difficult task due to frequent insertions and deletions of amino acids. Using comparative modeling, 80% of main-chain backbone atoms in the structure of target, whose sequence similarity lies in the range of 40–30%, can be predicted having rmsd error up to 3.5 Å [15–19], remaining 20% residues are modelled with high percentage of error specially in the insertions and deletions regions. Major problems associated with the modeling of protein having sequence similarity less than 40% are not only in obtaining correct sequence alignment but also in the modeling of loop and in placement of correct side-chain pose. Lastly, if the sequence similarity is less than 30%, then main problem in the modeling is the identification of the homologue structures, and alignment of target-template sequence becomes too difficult. Incomplete information about the structural

space in this region (*twilight zone*) implies the limited usage of the comparative modeling. Likewise, Comparative modeling is generally inhibited to confident homology detection region.

2.1.5. Comparative homology modeling and methodologies used by model building programs

The key step of comparative modeling is model building which involves explicit prediction of the atomic coordinates for the target structure from the equivalence residues present in pairwise alignment. Model building methods generally are of four types – 1) rigid body assembly [11, 20], 2) segment matching [21], 3) spatial restraints satisfaction [22] and 4) artificial evolution. These model building methods with their working principle and examples are described in the following paragraphs.

First one is *rigid body assembly*, as the name rigid body assembly, itself describes the working style of this methodology; It means small-small rigid bodies derived from the structurally conserved aligned regions of proteins are assembled to model the complete structure of target protein. In this methodology, accuracy depends on template and alignment quality. This methodology is used by the programs SWISS-MODEL and 3D-JIGSAW. Second one is *segment-matching* approach, which uses a subset of atomic positions, obtained from the matching segments during alignment in a database of experimentally known protein structures. These short segments are selected and used as building blocks in the construction of all atom models. Selection criteria for these segments are done using either energy or geometry rules or by permutation of these both criteria. This approach is widely used by many programs such as SegMod/ENCAD, *etc.* Third methodology is *spatial restraints satisfaction*, and in this methodology, the model is constructed using a set of restraints derived from the alignment and then minimization was done to decrease the violations of these restraints. Idea behind this methodology is that evolutionary conserved residues exhibit similar structural features and these structural features are imposed in the form of restraints. These restraints are generally supplemented with generic stereochemical restraints from molecular mechanics force fields. Restraint-based modeling methodology is better and most flexible as different types of restraints and constraints are easily incorporated in it. Restraints can be derived from template structure or from experiments like NMR, fluorescence, and site direct mutagenesis. The difficult tasks noticed in model building using this methodology are prediction of

conformation suitable for side-chains and loop region. This approach is used in one of the most frequently used modeling programs, Modeller [22]. The fourth method is *artificial evolution*, in this method, the program start changing the template structure *via* mutation, insertion and deletion according to the query-template sequence alignment. After every modification the minimization is performed using simplified potential function to calculate the energy, and change having minimum energy is selected for the next step and this process is repeated until the query sequence is completely modeled. The artificial evolution methodology of model building has been implemented in the NEST program.

2.1.6. Refinement of model

The refinement of initial model predicted by comparative modeling normally involves an effective sampling strategy for the conformational space and accurate energy function to identify near native structure. Primary goals of the model refinement methods are tuning of alignment, modeling of loops and optimal rotamer selection for the side chains. In other words, it involves simultaneous optimization of the predicted conformations for the non-conserved regions along with the physically adjacent regions. This is generally achieved by the molecular dynamics techniques using a correct force field and appropriate environment, it leads protein model to its native conformation. However, in the CASP experiment, it is concluded that the conformational space explored in the MD simulation is fairly local to the initial conformation [23]. There are four possible reasons for this problem: inadequate sampling of the possible conformations, the ruggedness of the potential energy landscape, having not enough accuracy in the inter-atomic forces description, and too short simulation times [23, 24]. To solve the above mentioned problems, optimization techniques perform the limited conformational sampling with a detailed force field, and then more extensive sampling with a simplified force field [25]. However, these approaches also found to be ineffective, due to which a refinement method using only MD simulation is useful for inadequate structural arrangements [26].

The refinement process generally starts with energy minimization step using suitable force field and for the next step, various techniques such as molecular dynamics, Monte Carlo and genetic algorithm based sampling can be applied. Refinement is generally focused on those regions of model which are probable to have errors, while remaining parts of the structures are allowed to relax in a physically realistic all atom force field. In this way,

refinement process significantly improves the quality of models by improving the backbone conformations as well as the placement of core side chains. Alignment accuracy strongly depends on the percentage of sequence similarity. The misalignment during the model construction results into the errors which are very hard to remove at the later stages of the modeling, *i.e.*, refinement [27].

The successful development of reproducible and reliable refinement methods, in which consistently improvement was done by shifting the coordinates closer to the native state, is one of the serious challenges in the field of structural bioinformatics. The activity of proteins in maximum cases involves the conformational changes, and understanding of molecular mechanism behind this requires a complete explanation of different functional states dynamically explored by the structure. The classic examples of this problem include conformational changes upon substrate bindings, excited states at the intermediate step in reaction cycles, transport activities, *etc.* However, these various states cannot be observed directly using high resolution experiments, and these states are generally characterized by the low resolution experiments, such as chemical shift changes, FRET, SAXS, or limited electron density in X-ray crystallography [28–30]. The Modeling and simulation is identified to play an essential role in exploring these alternative conformations and in explaining the dynamics of the structure.

2.1.7. Application of homology models in structure-based drug discovery

The rational drug development mainly depends on the structure-based strategies for identifying the effective as well as selective low molecular weight compounds, and the entire process in turn depends upon the quality and accuracy of the homology model. Homology models finds widespread applicability in the structure based virtual screening and has been utilized in various retrospective analyses on a variety of different targets [31–34]. Similar to the homology modeling technique, the efforts from the research community are needed to assess the correctness and consistency of modelled protein-ligand, protein-DNA, and protein-protein interactions [35–37].

Ripphausen and co-workers have analyzed 322 potential virtual screening operations, and out of total 322, in the 73 virtual screening studies, the homology models were successfully utilized [38]. And interestingly, potency of the hits recognized using the homology models was found to be higher in comparison to the hits identified through the X-

ray structures docking. In structure-based drug discovery, it is noticed that experimental structures will not necessarily provide better results than the models. In the recent study, a model of protein CYP2D6 was constructed based on the experimental substrate-bound CYP2C5 structure and at the same time, the experimental structure of apo-CYP2D6 also became available. So both the *in silico*, as well as experimental structures were taken as receptors in the docking calculations, irrespective of the fact that *in silico*, as well as experimental structures were in good agreement and the performance of the model was superior than the experimental structure in the docking calculations. The better performance can be credited to structural differences in the substrate binding sites and to the correctly described conformational changes taking place upon the substrate binding. Based on the above observations it can be concluded that the modeling techniques can play a critical role in searching the optimal receptor conformations. Application of the modeling is not only limited to the development of new drugs, but also it plays important role in various other biochemical processes, for example, the effect of amino acids mutations in the binding process. However, conveying the taken assumptions of the computational technique, and assessing the required accuracy of a model will be decisive to make these techniques helpful for the non-expert researchers. Understanding the limitations of a model is very crucial to decide which interpretations and conclusions can be supported. One of possible solutions in this respect will be to develop an open environment for sharing of data, models, and algorithms which will allow us to continuously and collaboratively refine our current model.

2.1.8. Limitations of recent advanced homology modeling techniques

The structural properties of proteins are very complex to understand as each protein is made-up of unique three-dimensional structure. As sometimes the small changes in the protein sequence result in strong effects on its structural properties and experimental determination of protein structures is a difficult process. Therefore, computational modeling of protein structure fascinated large interest in the field of structural bioinformatics to complement the efforts of experimental structural biology in characterizing the protein universe. In the following paragraphs, some applications of structure modeling and prediction techniques their limitations in communicating model information are presented.

Nowadays, the template based homology modeling techniques are so well established that it can be considered as the complement to the experimental techniques. Importantly, the

integration of the data obtained from the low resolution experimental technique with the computational modeling methods allows the study of routine as well as large complex systems with high accuracy. For example, the use of high ambiguity driven protein-protein docking (Haddock), it is different from the *ab initio* docking methods because it encodes the experimental information as the ambiguous interaction restraints (AIRs) to drive the docking process [39]. Experimental information from the low resolution experiments like mutation, SAXs, cryoEM, biophysical, biochemical and NOE, *etc.*, are converted into the AIRs to drive docking process to achieve results with the high accuracy. Haddock can be used to solve a large class of modeling problems including protein–protein, protein–nucleic acids and protein–ligand complexes. However, the incomplete knowledge of the computational modeling techniques is still facing so many challenges which hamper its efficient utilization by the non-expert researchers. The key problems associated with the modeling are difficulty in conveying the fundamental assumptions of a given computational technique, and the expected correctness and structural inconsistency of a given model. However, both these aspects are very essential to recognize the limitations of a model, and to decide that which analysis and interpretations can be carried out.

In the 1987, researchers Box and Draper stated that all models are wrong, but some are useful [40]. The different applications of macromolecule structure need different levels of accuracy and resolution, while atomistic modeling requires highly accurate set of coordinates. Models of lower resolution are also suitable for various types of studies including rational drug design, epitope mapping, site directed mutagenesis and supporting experimental structure determination [41]. To judge the model applicability, it is very crucial to know its accuracy as well as quality.

The model assessment techniques are retrospective experiments such as CASP or CAMEO allowed comparing different methods on the same data set and in this way help in finding the area which need further development or improvement in the methods. However, the differences in the accuracy between the same protein targets are generally small for the same protein target and differences between the easy and difficult protein targets are generally high. Therefore, estimation of the local error for the atomic coordinates generated by a model is essential to judge its applicability for a specific biomedical question. This is one of the major limitations in the success of current modeling techniques, and it hinders the widespread use of models in the area of biomedical research. To avoid this problem, a number of tools for

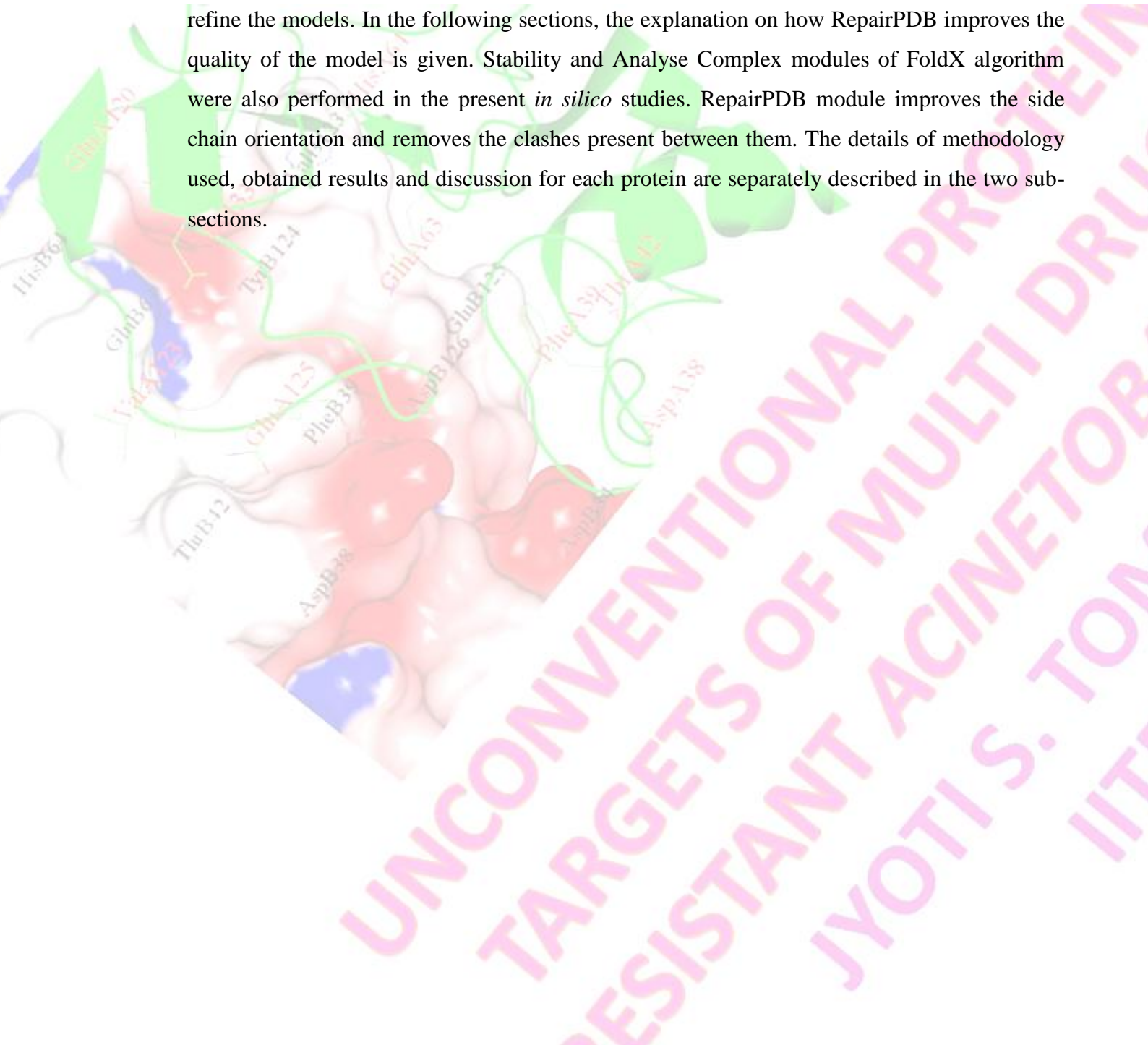
the model validation and quality assessment have been developed, which measures various structural features of a model such as correctness of stereochemical parameters (Procheck), or apply a combination of knowledge-based statistical measures obtained from the high-resolution structures. Validating the accuracy and reliability of a model is very important in order to estimate its applicability. The prediction of model suitability for a given application requires two things first one is obviously the access to the model coordinates and second one is the information about the procedures and necessary assumptions made during the model generation. As the coordinates derived by theoretical modeling cannot be deposited in the PDB due to which many manuscripts reporting results of *in silico* modeling and simulations are published without submitting the model in the molecular modeling database available, or without making the model accessible. This is highly problematic for the reviewer of the manuscript to judge if the experiment is reproducible and given conclusions are vindicated. In order to improve such situation, a public archive of theoretical models (<http://modelarchive.org>) is currently being established as a part of the Protein Model Portal [42]. Model archive provides a unique accession code (DOI) for each of the deposited model, which can be given in the corresponding manuscripts so that reviewers can verify the results. Along with the actual model coordinates, archiving of models also includes the necessary details about the assumptions, parameters and constraints applied in the simulation.

2.1.9. How our optimized methodology is useful in the protein modeling?

As we discussed in the previous sections, homology modeling is a depiction of the residues of similar environment at topologically corresponding positions in the template proteins. In the absence of experimental structure, model constructed on the basis of a known 3D structure of a homologous protein is an only reliable method to obtain the information about the structure. The Knowledge of proteins structure provides valuable insights into the molecular basis of their action. In this chapter, we performed the homology modeling of two proteins TAG and Hpa2 from the organism *A. baumannii*. For the model construction we used one of the well known programs, *i.e.*, Modeller9v12. After model construction the best model was selected on the basis of the DOPE score and model having better structural parameters. Then energy minimization was performed to improve the model quality.

It is well known from the literature that modeller performance is excellent in most of the aspects. However, performance is found to be poor in terms of placement of the side

chains. Since we used modeller to construct the model for both of proteins, and keeping the problem of the side chain quality in mind, we searched for the program which can solve this problem. Finally we found that it is suitable to use the RepairPDB module of FoldX to further refine the models. In the following sections, the explanation on how RepairPDB improves the quality of the model is given. Stability and Analyse Complex modules of FoldX algorithm were also performed in the present *in silico* studies. RepairPDB module improves the side chain orientation and removes the clashes present between them. The details of methodology used, obtained results and discussion for each protein are separately described in the two sub-sections.



Section-I

The genus *Acinetobacter*, especially the species *A. baumannii* has gained lot of attention as one of the critically emerging bacterial pathogens [43, 44]. The successful emergence of *A. baumannii* is a major alarm for human health because it has enormous capacity of rapid transformation and showed resistance to almost all currently available antimicrobial agents [45]. The organism causes outbreak of health-care associated infections including bacteraemia, UTI, meningitis, pneumonia, endocarditis, keratitis, osteomyelitis and peritonitis [46, 47]. Resistance mechanisms for *Acinetobacter* species are divided into three categories: (i) antibiotics-inactivating enzymes, (ii) abridged access to antibacterial targets and (iii) resistance *via* mutations that increases expression of operons and intrinsic genes [48]. Owing to the lack of good therapeutic options as well as developing resistance toward majorly available antibiotics, it is difficult to cure the infections caused by *Acinetobacter*. Consequently, the innovation of new antibacterial scaffolds would be a better strategy to keep multidrug-resistant infections at bay. Hence the discovery of novel targets from metabolically important pathways may turn to decisive proposal. The idea behind the present study led us to the search of novel targets in *A. baumannii* which concluded at Hpa2, a member of GNAT superfamily of HATs due to the vindication noticed in literature as described below.

Histone acetyltransferases (HATs) are enzymes known to transfer an acetyl group from acetyl-coenzyme A (acetyl-CoA) to ϵ -amino group of conserved lysine amino acids on histone *proteins* or to small molecules like polyamines [49]. HATs are classified on the basis of sequence homology and structural features. Superfamily Gcn5-related *N*-acetyl transferases (GNAT) illustrate less sequence homology irrespective of remarkably conserved fold of catalytic domain [50]. This superfamily includes members like Gcn5, PCAF, Hat1, Elp3, Hpa2, Hpa3, ATF-2, and Nut1, which act on a range of substrates such as lysine residues, aminoglycoside antibiotics, spermine/spermidine or arylalkylamines crucial for the cellular functions [51–53]. During literature inspection, we noticed two interesting things; firstly, study performed by Blanchard and co-workers on *S. enteric* revealed that AAC6, an enzyme is capable of regioselective *N*-acetylation of antibiotics. AAC6 enzyme exhibits dimeric structure and its structure is very similar to *Hpa2*-encoded histone acetyltransferases of *S. cerevisiae* [54]. Secondly, the biochemical characterization of Hpa2 and Hpa3 proteins illustrates that both of these enzymes have capacity to acetylate histone and polyamines [55].

It is well documented that polyamines are imperative modulators of a variety of ion channels [56], gene expressions and protein activity [57–59]. Also, addition of polyamines to the growth medium of *P. aeruginosa*, a MDR bacterium, was found to increase the MIC values of multiple antibiotics. This suggested the possibility of a molecular mechanism for polyamine-mediated antibiotic resistance [60–61]. Based on the evidences obtained from the literature (*vide supra*) on acetylation of polyamines by Hpa2 and structural similarity with enzyme AAC6, it can be hypothesized that Hpa2 might be engaged with polyamine dependent or independent antibiotic resistance mechanism. Moreover, the exact molecular function of the Hpa2 in the cell is not well characterized, due to which, three-dimensional structure was predicted computationally to shed more light on the usefulness of structural information ranging from precise family and functional assignment to structure-based drug screening.

Generation of accurate protein model from its less conserved sequence information is a very critical step and for this we have developed a new protocol that can generate highly accurate protein models in terms of optimal backbone dihedral angles ψ/ϕ , side-chain orientation, and binding sites of substrate. To obtain accurate protein models, we have developed an efficient, cost and time effective novel protocol for residue-level protein structure refinement. In the present study, steps pooled with traditional homology modeling are multiple sequence alignment and its manual adjustment, loop refinement, side-chain remodeling *via*, the optimal combination of side-chain rotamers and finally improving the energetics of model using the *FoldX commands*.

2.2. Methods

The workflow for constructing monomeric and dimeric Hpa2 model is depicted in Chart 1. The protocol for generating monomeric tertiary structure of Hpa2 is discussed below.

2.2.1. Methodology for generating monomeric tertiary structure of Hpa2

For less homologous sequences *i.e.*, less than 35%, generation and optimisation of optimum quality model *via* conventional method is a difficult task or nearly impossible. To circumvent these problems, conventional homology modeling of Hpa2 was pooled with a few commands of *FoldX* algorithm to achieve high accuracy in structural parameters. The amino acid sequence of Hpa2 retrieved from the NCBI database with accession number ACC58416 was BLAST [62] against protein data bank <http://www.rcsb.org> to obtain the homologous template

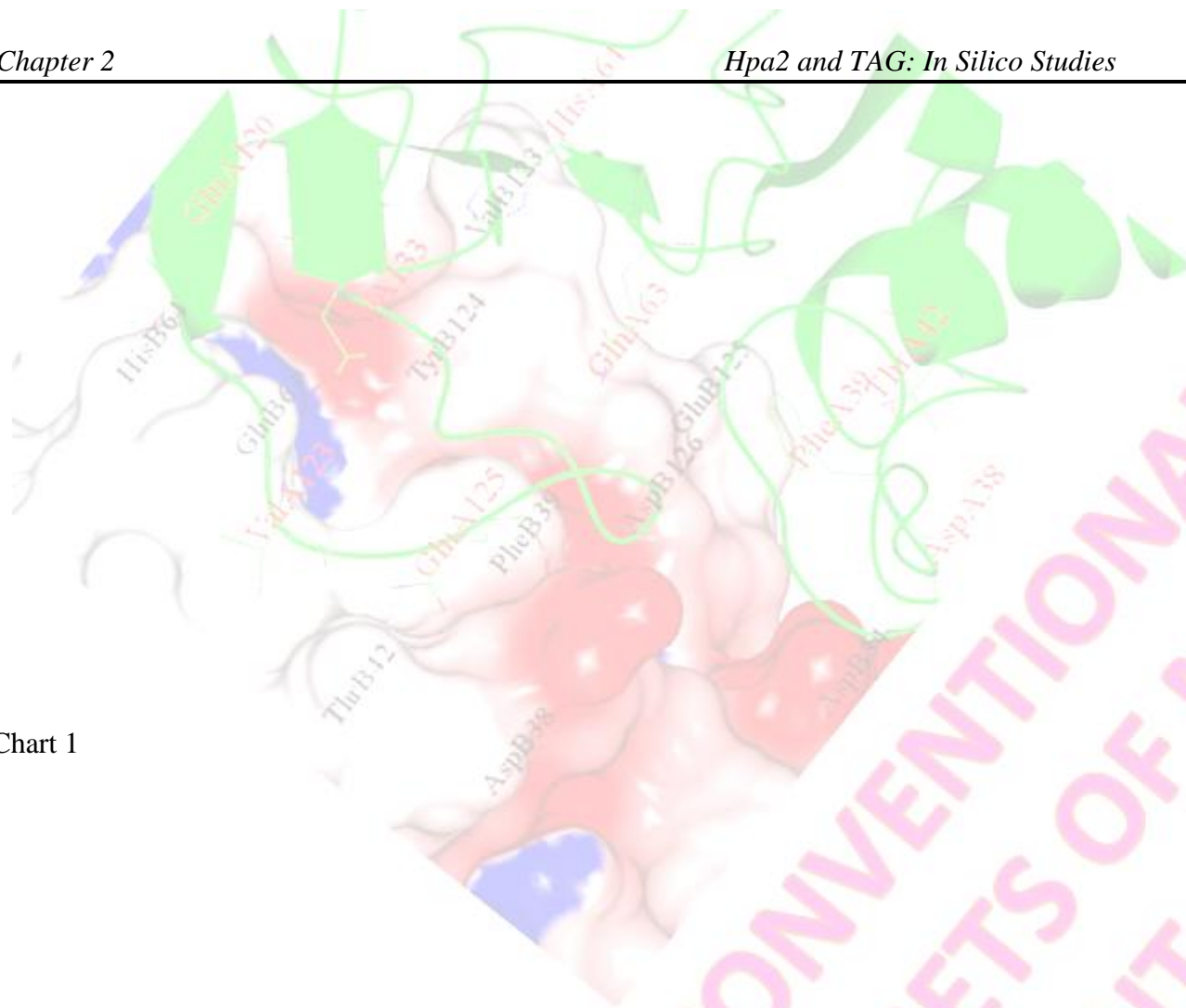


Chart 1

UNCONVENTIONAL PROTEIN
TARGETS OF MULTI DRUG
RESISTANT ACINETOBACTER
JYOTI S. TOMAR
IITR

structures for the homology modeling. The pairwise sequence alignment with homologues templates was performed using T-Coffee [63] and manual adjustment was done to improve the quality of alignment [64]. The monomeric model of Hpa2 was prepared using SWISS-MODEL which incorporates model construction and minimization steps [65]. Then optimization step was completed by subjecting the model to *RepairPDB* module in FoldX command [66]. For accessing the model quality, two types of evaluations were performed; First one is *internal evaluation* using PROCHECK [67] and Verify3D [68] for self-consistency checks to see whether model satisfies the restraints used to calculate it. Second one, PROSA Z-score for *external evaluation*, utilises the information used in the calculation of the model [69]. Finally, energy calculation was performed using *stability* command of FoldX algorithm. It provides Gibbs free energy of folding composed of various energy terms, which is useful for the model appraisal. Eventually, the model is checked for its consistency with experimental observations such as site-directed mutagenesis and ligand binding. The generated Hpa2 monomer structure is now available for the structural analysis, virtual screening and protein–protein docking experiments.

2.2.2. Methodology for designing of dimeric Hpa2 structure

Prior to developing dimeric Hpa2 model during this study, we decided to reconstruct the experimentally available dimeric structures for acquiring optimal theoretical parameters. For this process, homologues template dimer structures present in the pdb were retrieved. *Analyse complex* command of FoldX algorithm was then used to determine different energy terms along with the complete list of residues involved in the interaction at the interface, which were further verified manually using the pymol plugin [70]. After that, heteroatoms were removed and modified amino acids mutated to normal amino acids with the help of *pymol-wizard* mutagenesis since FoldX does not recognise all types of modified amino acids. The interface residues that make the direct contact are considered as active residues and remaining interface residues were further categorized as passive residues in Haddock [39]. In this data-driven docking study, the experimental information entered in the form of active and passive residues, is further converted by the program into ambiguous interaction restraints (AIRs) with default value of restrain partitions ($n = 2$) to perform the docking. The AIR is defined between each active residue of one protein and all the active and passive residues of the partner protein. The Haddock uses Crystallographic and NMR System (CNS) as its structure

calculation engine [71]. It generates the topology of the input structure just before the docking simulation. Afterwards, the simulation process starts and this consists of three sub-steps *viz.*, energy minimization, semi-flexible simulated annealing to refine the interacting surface and then explicit solvent model which is used for the final refinement. The structures were ranked after completion of every step and best solution is chosen for the next stage. Successful completion of docking run produces the clusters numbered according to cluster size and these clusters are sorted by Haddock score. The best docked dimeric structures are further taken for rmsd value calculation to ensure that applied protocol is suitable for this type of biomolecular system. Then the same protocol was followed for the Hpa2 dimer formation. For this, active and passive interface residues were identified through the sequence alignment with homologous dimeric templates (1XEB and 1QSM). Then conserved residues present at the interface are taken as active and residues surrounding them are considered as the passive residues. These residues are used as AIRs for the generation of dimeric Hpa2 by the Haddock. Then the structure was optimized and validated in a way similar to that of the monomer structure. Finally, the interaction energy was recalculated using the *Stability* command of FoldX algorithm. The different energy terms used in the FoldX provide the rough estimate of the importance of different forces in the formation of the complexes analyzed here. The formed Hpa2 dimer structure was used for the structural and interface analysis and virtual screening.

2.2.3. Docking of acetyl-CoA-based compounds

For inhibitor designing as well as to characterize the enzyme action, it is vital to have information about the nature of residues making the active site and forces involved in the optimal binding. For this purpose, Vina wizard module of PyRx, which takes receptor as a rigid structure to carry out the virtual screening of large number of possible ligands, was considered [31, 72]. Herein, a library of 670 compounds, having at least 60% similarity with substrate acetyl-CoA, was generated from ZINC database <http://zinc.docking.org>. For the geometry optimization, we used the Steepest descent method for the first 1000 steps and then switched to the Conjugate gradient method for the next 2000 steps. The low energy conformers of each compound were converted to pdbqt file and then subjected to virtual screening [73]. The docking of polyamines and aminoglycoside class of antibiotics were also performed on both monomeric as well as dimeric model to test the hypothesis made at the beginning about the antibiotic resistance properties of the enzyme Hpa2.

2.3. Results and Discussion

The ever-rising gap between genome sequences and experimental structures of proteins is very challenging task for the structure-based drug discovery efforts. Although computational structure prediction methods provide a cost and time effective alternative but developing enough accurate models using the remotely homologous templates is still very challenging. The developed protocol herein is quite easy, efficient in terms of time as well as computation cost and sufficiently accurate for the modeling of proteins having distant homology. Furthermore, the protocol can be used for the creation of dimeric model successfully.

2.3.1. Alignment of amino acid sequences and secondary structure prediction

The templates used for the homology model construction were retrieved from the pdb database using the BLAST search. The search shows that Hpa2 of *A. baumannii* shares 35% sequence identity with GNAT superfamily protein Yjcf from *B. subtilis* (PDB: 1Q2Y). Furthermore, it has 29% and 33% sequence identity with GNAT superfamily members *L. plantarum* (PDB: 3EFA) and *P. aeruginosa* (PDB: 1XEB), respectively (Table 1). Since the Hpa2 of *A. baumannii* showed highest percentage of sequence identity and query cover with Yjcf protein from *B. subtilis*, it was considered as an appropriate template for homology

Table 1. Templates from BLAST against PDB with Hpa2 of *A. baumannii*.

PDB id.	Organism	Title	Sequence identity	E-value
1Q2Y	<i>B. subtilis</i>	Chain A, Crystal Structure of the Protein Yjcf	35%	4e ₋₁₇
3EFA	<i>L. plantarum</i>	Chain A, Crystal Structure of Putative <i>N</i> -acetyltransferase	29%	7e ₋₁₁
1XEB	<i>P. aeruginosa</i>	Chain A, Crystal Structure of an Acyl-CoA <i>N</i> -Acyltransferase	33%	5e ₋₀₇
2JDC	<i>B. licheniformis</i>	Chain A, Glyphosate <i>N</i> -acetyltransferase Bound to Oxidized CoA and Sulfate	29%	2e ₋₀₆
2BSW	Synthetic	Chain A, Crystal Structure of A Glyphosate- <i>N</i> -Acetyltransferase Obtained by DNA Shuffling	28%	4e ₋₀₆
1GHE	<i>P. syringae</i>	Chain A, Crystal Structure of Tabtoxin Resistance Protein Complexed with an Acyl Coenzyme A	28%	0.004
1J4J	<i>P. syringae</i>	Chain A, Crystal Structure of Tabtoxin Resistance Protein (Form II) Complexed with an Acyl Coenzyme A	28%	0.004
2DXQ	<i>A. tumefaciens</i>	Chain A, Putative Acetyltransferase	31%	0.005

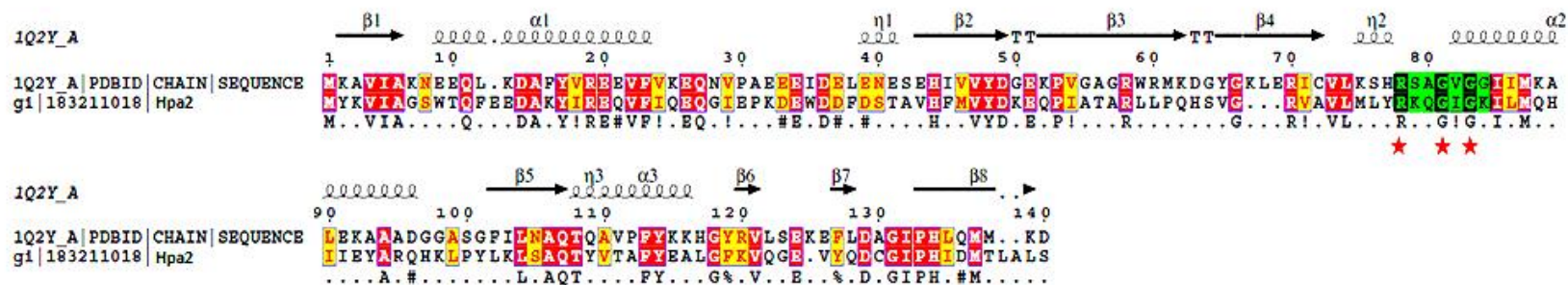


Figure 2. Sequence alignment and secondary structure prediction of the enzymes. Hpa2 (Accession ACC58416) and template (pdb id: 1Q2Y_A)
* Indicates the conserved GNAT motif residues.

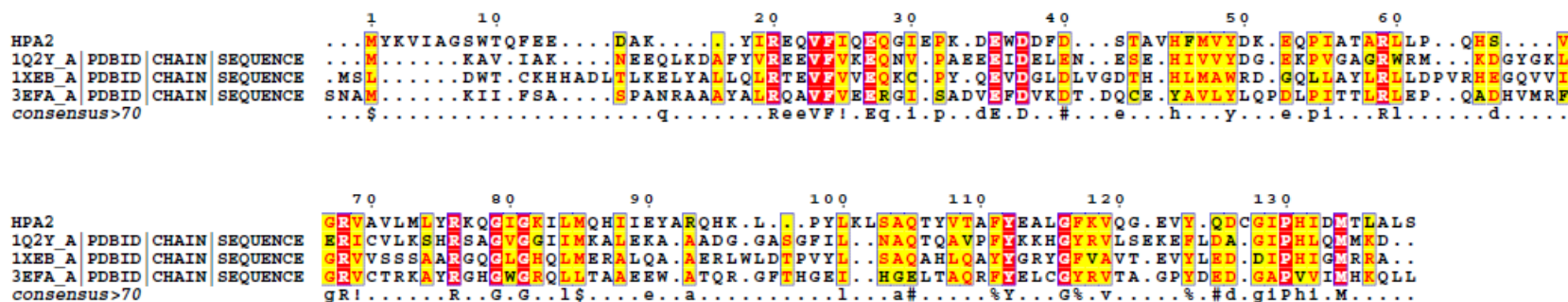


Figure 3. Multiple sequence alignment of Hpa2 sequence with its templates having sequence similarity up to 29%.

modeling. Secondary structure prediction of Hpa2 was done using ESript using Yjcf protein of *B. subtilis*; (PDB: 1Q2Y, secondary structure as reference). Inspection of [Figure 2](#) corroborates the well known statement that member of the GNAT superfamily proclaims conserved catalytic domain fold among them, irrespective of low percentage homology. The residues present at the secondary structures $\alpha 1$, $\beta 4$, $\beta 4/\alpha 2$, $\alpha 2$, $\beta 5$ and $\alpha 3$ are found to be relatively more conserved. The multiple sequence alignment of homologous GNAT family members again verifies the well known concept about the sequence similarity of GNAT family, *i.e.*, presence of variable amino-acid sequences between the closest related homologues, however, retaining high structural homology ([Figure 3](#)).

2.3.2. Hpa2 monomer model construction, optimization and quality assessment

After the construction, monomeric Hpa2 model was subjected for further refinement, using the *RepairPDB* function of FoldX algorithm, for quick removal of incorrect torsion angle van der Waals clashes and residues having high energy by exploring different combinations of rotamers of lowest possible energetics. The *RepairPDB* function improves the structural parameters as well as energetics efficiently within few minutes. Finally, *PrintNetworks*, *Alascan* and *QualityAssessment* commands were performed to identify and to rectify the apparent problems detected in the model structure.

Quality assessment of model is very crucial to judge whether this model would be suitable or not to tackle a particular problem, model quality was validated with the help of PROCHECK, in which the statistical analysis of dihedral angles shows that 96.2% residues are scattered in core and allowed regions and remaining 3.8% residues are present in the generously allowed (2.5%) and disallowed regions (1.3%). Furthermore, the G-factor estimation was also performed, which calculates covalent geometry and torsion angles to determine the degree of normality for the given structure. The value of G-factor for dihedral torsion angles is 0.01 and that for covalent forces is 0.46 and the overall G-factor score is 0.19. Residues K51 and Y99 of disallowed regions are present far away from the binding pocket of enzyme ([Figure 4](#)), convincing that the bad torsion angles of these residues do not affect the enzyme activity. Varify3D result shows that 80.4% of residues have averaged 3D-1D score of > 0.2 , which means 80.4% of Hpa2 amino acid residues are assigned a correct structural class-based on its location (α or β or turn) and environment (hydrophobic or hydrophilic). For low sequence identity cases such as ours, misalignment and insertion in

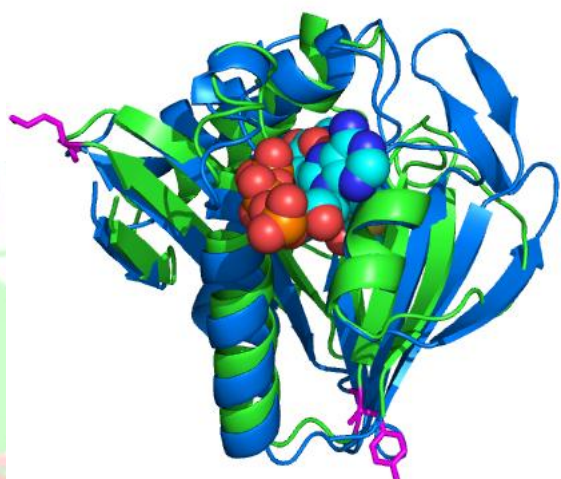


Figure 4. Model (green coloured) superimposed on template 1GHE (blue coloured) complexed with acetyl-CoA (sphere representation). Amino acid residues present in disallowed region (stick representation, magenta colour) are present far away from the binding pocket.

query sequence may take place which negatively affect the model quality and the effect may be seen as rise in main chain rms error by 1.5 Å for about 80% of residues. To counter such problems in this study, PROSA Z-score value (−5.75) was calculated for Hpa2 model which was comparable (−5.60) with the template 1Q2Y (Table 2).

Table 2. Summary of structural validation programs and parameters values for Hpa2 monomer model and dimer model with their respective templates (pdb id: 1Q2Y, 1XEB).

Server	Parameters	Repaired Hpa2 Monomer	Template 1Q2Y	Dimeric Model	Template 1XEB
PROCHECK	Residues in core, Allowed	85.4%, 10.7%	90.1%, 9.9%	81.8%, 15.7%	85.6%, 14.4%
	Generously allowed and	2.5%	0.0%	1.7%	0.0%
	Disallowed regions	1.3%	0.0%	0.8%	0.0%
PROSA	PROSA Z-score	−5.75	−5.60	−5.68	−4.58
VERIFY3D	Residues having averaged 3D-1D score > 0.2	80.01%	80.43%	80.24%	80.30%
ERRAT	Overall quality factor	81.29%	88.97%	96.87%	93.39%

This symbolizes that the template chosen is correct as well as the query-template sequence alignment was sufficiently correct. Analysis of energy terms of the model structure using *stability* of FoldX showed that the stabilizing energy terms are well compensated by the destabilizing energy terms. Superimposition of Hpa2 with its templates for calculating the rmsd showed a compact perfectly fitted secondary structure (Figure 5A). The overall fold of Hpa2 monomer is quite similar with template 1Q2Y (rmsd = 0.45); however, with 1XEB

(rmsd = 2.02) secondary structures are also well fitting with some differences in the loop regions of the structure. The results of the above mentioned structure validation programs, calculated rmsd values and energy of the model (id: y853suf5milsp99eb; Model Archive doi: ma-arf2p) attest that Hpa2 model quality was satisfactory and sufficiently accurate for further study.

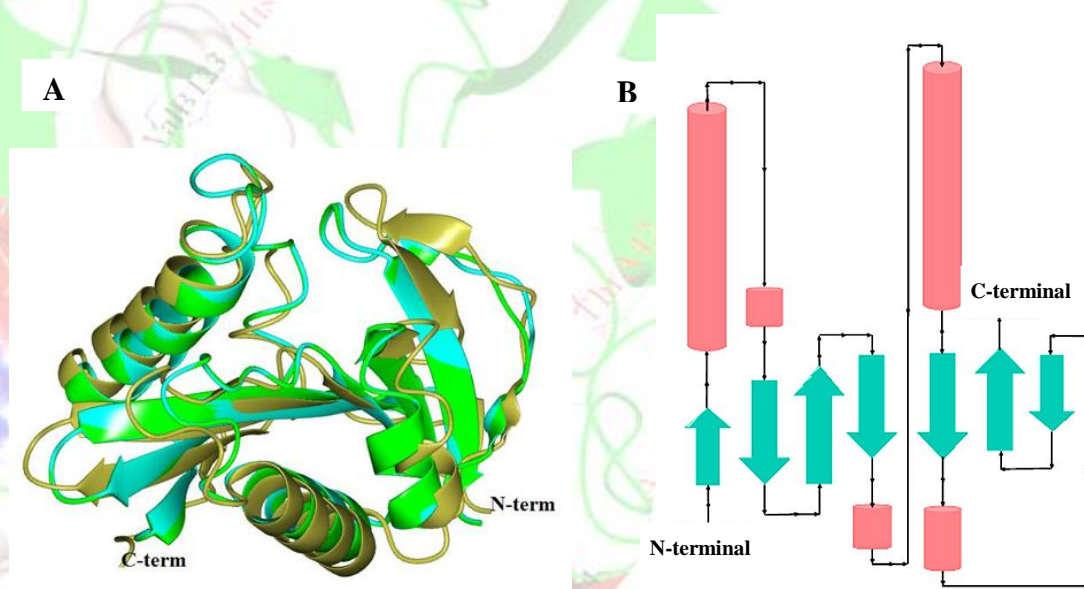


Figure 5. (A) Structural alignment of the enzymes; predicted model (green) and templates (pdb id: 1Q2Y_A, rmsd = 0.45, cyan) (pdb id: 1XEB_A, rmsd = 2.02, gold). (B) Topology diagram for the Hpa2 model, there are a total of seven β -sheets and five α -helices (β -sheet 1; α -helices 1 and 2; β -sheets 2, 3 and 4; α -helices 3 and 4; β -sheet 5; α -helix 5; β -sheets 6 and 7).

2.3.3. Comparative structural analysis of monomeric Hpa2 and its relatives

Hpa2 is consisting of the common GNAT fold, which is made up of seven β -sheets and five α -helices in the form of secondary structure. The 139-amino acid Hpa2 exhibits the connectivity $\beta_1, \alpha_1, \alpha_2, \beta_2, \beta_3, \beta_4, \alpha_3, \alpha_4, \beta_5, \alpha_5, \beta_6$ and β_7 (Figure 5B) and most of the β -strands are packed in antiparallel fashion except β_4 and β_5 which are packed as parallel β -strands. The number, type and connectivity of secondary structures in the Hpa2 model are identical with the tabtoxin resistance protein (TTR) except the α_3 helix, which is present after $\beta_2\beta_3\beta_4$ in Hpa2 and before $\beta_2\beta_3\beta_4$ in the TTR [74]. The enzyme TTR catalyzes the acetylation of tabtoxin, which is precursor to the antibiotic tabtoxinine β -lactam (TL) produced by *P.*

syringae. This structural similarity might be allied with the antibiotic modifying properties of the Hpa2. Similar to yeast Hpa2, the model includes conserved $Q\backslash R-X-X-G-X-G\backslash A$ segment [75], (coloured green in Figure 2) is known to be involved in the main chain contacts and it is well characterized that mutation of even a single residue in this motif causes drastic reduction in the enzyme activity [76, 77]. Structural analysis performed here deputed that similar to other GNAT structures, Hpa2 possesses a conserved β -bulge at the position which is known to make the V-like appearance.

2.3.4. Regeneration of homologous experimental dimeric structures

Dimerization of monomer unit might create novel binding sites such as dimeric interface or it could extend the number of active binding faces. In the case of yeast Hpa2, dimerization is found to be related with the increase in the number of active site. In the present study, we generated the dimeric model to characterize the interface and the binding sites. For the generation of dimer Hpa2, we need optimized parameters and to accomplish this goal, reproduction of the homologues experimental dimer template structures (1XEB, 1J4J, 1GHE and 2DXQ) having identity up to 28% and query cover up to 51% were made. The interfacial residues of experimental homologues dimer are given in Table 3. The docked 1XEB structure (chains A and C) was found to superimpose well on experimental structure (Figure 6A). Redocked experimental dimer shows the rmsd values 3.02 Å, 2.35 Å, 3.23 Å and 9.3 Å for the pdb 1XEB, 1J4J, 1GHE and 2DXQ, respectively. Redocked dimer structures were successfully reproduced by the data-driven docking except for the 2DXQ, since 2DXQ structure is assembled by extensive interactions in secondary structure and approximately 33% of its residues are involved in the construction of dimer interface. Approximately, 397 non-bonded interactions and 34 H-bonds are found to be formed at the interface. The homodimeric structure of 2DXQ is formed *via* extensively integrated secondary structures, due to which it is a difficult target to be reproduced theoretically. On average, 9–14 residues of each chain interact to craft the interface in each of 1XEB, 1J4J and 1GHE dimeric structures (Figure 7). A total of about 30 non-bonded contacts were found for 1J4J and 1GHE while the value is notably higher (~45) for 1XEB. However, disulfide bonds, salt bridges were absent at the interface. The successful regeneration of experimental complexes signifies the suitability of the optimized parameters and the program Haddock.

Table 3. PDB files, residues involved in the formation of experimental dimer interface and rmsd values for redocked experimental dimers calculated using experimental structure as reference.

PDB/ MODEL	Interface residues	RMSD value
1XE-AC	A-Chain: H72, E73, G74, Q75, V77, Y114, V133, T134, Y137 and R148. C-Chain: V70, E73, G74, Q75, V77, Y114, V133, T134, and R148.	3.02
1J4J-AB	A-Chain: Q16, L17, Q121, D124, E125, E127, Q128, K182 and L184. B-Chain: Q16, L17, R18, T21, E23, H27, Y28, A79 and D81.	2.35
1GHE-AB	A-Chain: H30, G31, P100, S101, R103, R105 and V136. B-Chain: H19, Q23, D42, M43, Q44 and Y47.	3.23
2DXQ-AB	A-Chain: L16, Q27, D30, P31, L33, A38, V41, A44, M45, P49, L51, V71, P72, N73, L74, T75, R76, A77, A78, R79, Y81, F83, F111, C115, Y116, V118, M119, L121, T122, Y133, F138, V139, K142, T143, F145, Q146, I147, R148 and G149. B-Chain: L16, Q27, D30, P31, L33, A38, V41, G48, P49, P72, N73, L74, R76, A77, A78, R79, Y81, F83, N86, R104, Y116, K117, V118, M119, L120, L121, T122, Q140, V139, N141, K142, T143, F145, Q146, I147, and R148.	9.36
Dimer-Hpa2	A-Chain: D38, F39, Q63, H64, Q86, P98, Y99, G121, V123, Y124 and D133. B-Chain: D34, D38, F39, S41, T42, L61, P62, Q63, V123, Y124, Q125 and D126.	4.41

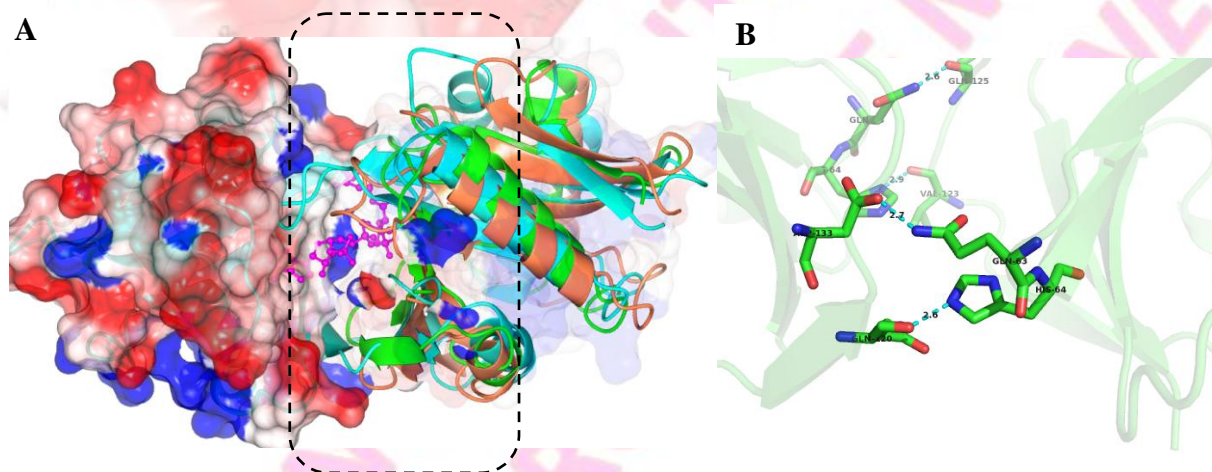


Figure 6. (A) The predicted dimeric docking conformation of model Hpa2 with acetyl-CoA (green) superimposed on the experimental structures 1XE-AC (cyan) and 1QSM-AB (Crimson red) for chain-A, electrostatic surface map (blue represents +ve charge, red for -ve charge and white for neutral charge) for the chain-B, the rmsd value is calculated keeping experimental structure 1XE-AC as reference (rmsd = 4.41Å). (B) Interface region of model Hpa2 showing the interacting residues from both the chains.

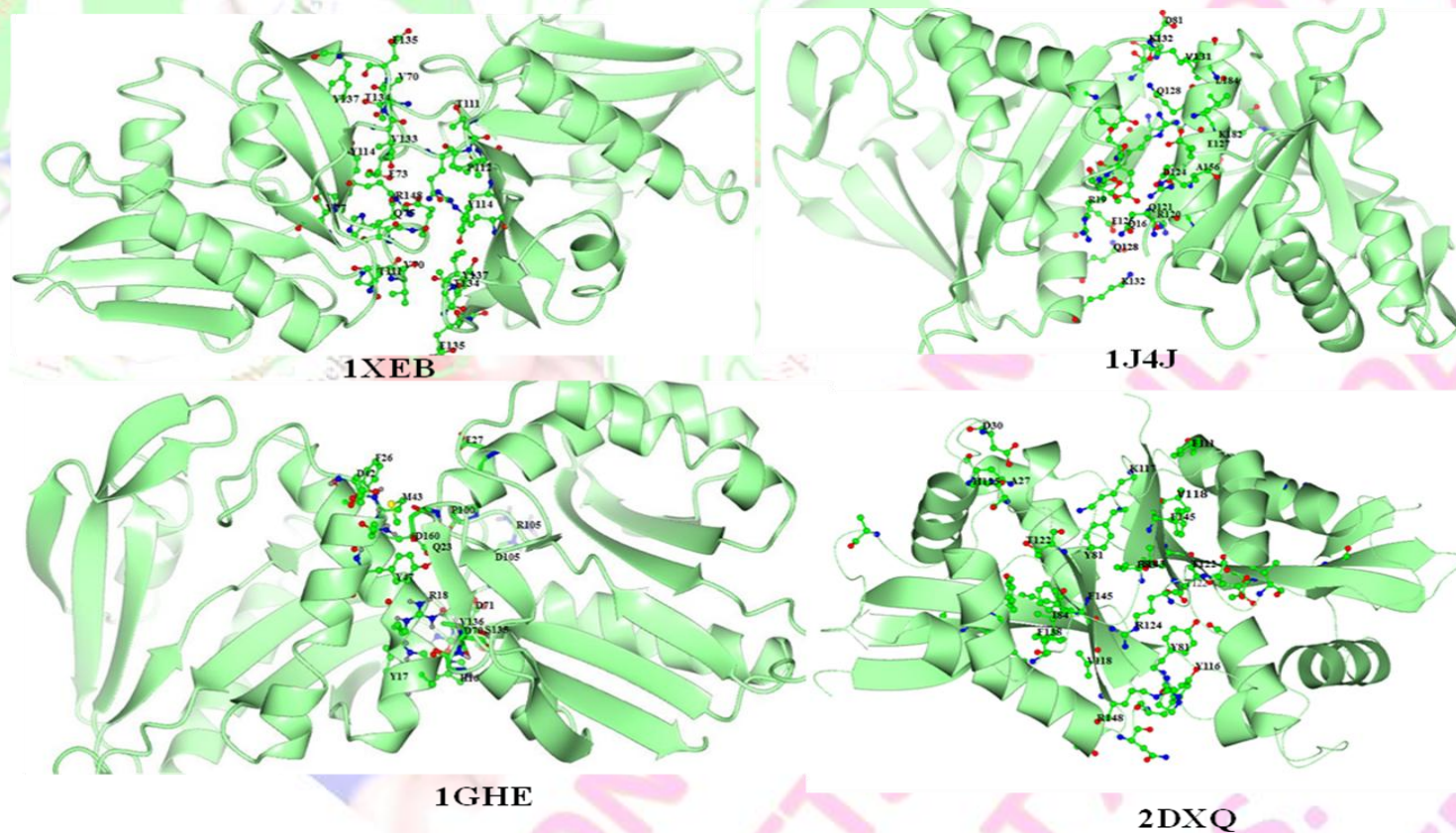


Figure 7. Experimental dimeric structures redocked using the data-driven molecular docking. The residues involved in the interface formation are labelled.

2.3.5. Construction, refinement and optimization of structural parameters and analysis of energetics of Hpa2 dimer model

Dimerization events in different protein families are known to generate huge functional diversity through different monomer combinations, dimerization event regulates diverse physical and biological outcomes. Monomeric Hpa2 model was subjected for data-driven docking using the protocol discussed in the *Methods section* (2.2). The alignment of query sequence with the 1XEB and 1QSM (*S. cerevisiae* Hpa2) revealed that eight conserved residues are present between 1XEB, 1QSM and query sequence and none of them contribute in the dimer interface formation. The 33% of conserved residues present between query and 1QSM participate at the interface. While approximately 9% of conserved residues found between 1XEB and query contribute at the dimer interface. These observations indicate that irrespective of less sequence similarity nature of Hpa2 dimer possesses similarity with yeast Hpa2, symbolizing that dimerization of Hpa2 to be a process initiated at prokaryotic Hpa2 and with evolution tight interactions acquired in the monomers of eukaryotic *S. cerevisiae*. So, prokaryotic Hpa2 could be considered as the evolutionary progenitor of eukaryotic Hpa2. The dimeric interface is approximately 1100 Å² in size and its size is consistent with homologous dimeric species 1XEB (1000 Å²). The rmsd value of Hpa2 keeping 1XEB_A as reference is high (2.02 Å), as result of low sequence homology. However, the rmsd value of dimeric model Hpa2 is 4.41 Å which is almost same as for the monomer signifying that the strategy followed for the dimeric Hpa2 construction is accurate (Figure 6A). Model quality assessment data (Table 2) establishes that the dimeric model quality was optimal and in fact better than that of monomer. Further improvement in the dimeric model may be attributed to the refinement steps in the work flow of Haddock run. The *Stability* analysis upon execution produces the Gibbs free energy of folding, fragmented into various energy terms (Table 4). For the dimeric model two extra energy terms ΔG_{kon} and ΔS_{tr} representing association constant and loss of translational entropy, respectively, were obtained. Analysis of *Stability* run on experimental as well as dimeric model Hpa2 provides a rough picture of various interactions responsible for the dimerization. The main contributions to the binding are from the ΔG_{vdw} and ΔG_{solvH} term. The 1QSM and 2DXQ dimers have the most favorable van der Waals interaction in all six dimers, which may be attributed to the better surface complementarity present at the interface. The residues present at dimeric Hpa2 interface were analyzed to understand the

nature of interface (Figure 6B). Four H-bonds were formed between Q63, H64, Q120 and D133 of chain A to V123, Q125, H64 and Q63 of chain B. The values of ΔG_{kon} for dimeric experimental complexes are given in Table 4, explicate that Hpa2 have more tendency of self association to form the dimer in comparison to 1XEB, 1J4J and 1GHE. The values of

association constants are almost double for the 1QSM and 2DXQ proteins and both of them are almost ten times more stable. The loss in entropy term upon dimerization is almost equal for all the dimers studied. The values of these energy terms again attest the dimeric Hpa2 model quality.

Table 4. Complex energies, and detailed energy contributions for the experimental dimers and for dimeric Hpa2 model (Output of *analyse complex* of FoldX). All values are given in kcal/mol.

PDB/Model	Monomer					Dimer			
	ΔG_{total}	ΔG_{H-bond}	ΔG_{vdw}	$\Delta G_{elect.}$	ΔG_{vdw} clashes	Entropy sum	ΔG_{solvH} + ΔG_{solvP}	$\Delta G_{kon.}$	$\Delta S_{tr.}$
1XEB	-03.06	-02.27	-08.97	-1.76	+0.27	+05.5	-0.8	-0.42	+2.34
1J4J	-03.68	-01.59	-07.47	-2.99	+1.03	+06.3	+7.0	-0.67	+2.31
1GHE	-04.19	-01.68	-06.76	-0.47	+1.02	+04.5	+2.5	-0.29	+2.34
2DXQ	-53.60	-32.98	-66.06	-4.37	+5.95	+56.5	-14.2	-1.08	+2.38
1QSM	-46.60	-44.30	-66.26	-5.57	+1.03	+58.4	+1.9	-1.57	+2.38
Model	-05.63	-04.68	-08.62	-0.39	+0.88	+06.6	-3.2	-0.83	+2.38

2.3.6. Virtual Screening against Hpa2 monomeric and dimeric structures

Virtual screening, particularly the structure-based screening, has become a cost-effective, reliable and obviously time saving technique for finding of new lead compounds. In this study, virtual screening approach was employed to discover new inhibitors for Hpa2 through screening substrate-based library, polyamines and antibiotics. The idea behind substrate-steered searching was to utilize the substrate binding information for the betterment in the shape complementarity as well as in binding forces. Herein, all the compounds selected for virtual screening exhibit binding affinity ranging from the -12 to -6 kcal/mol with Hpa2 proteins. The selected best ten compounds (Table 5) having high binding affinity with the monomeric as well as dimeric Hpa2 were further analyzed. Moreover, these compounds, when superimposed with both forms of Hpa2 to check the deviation in the binding pose between them, showed very low rms deviation indicating that the docked conformations of these compounds are almost same as those with the substrate acetyl-CoA. The superimposition of the Hpa2 model on experimental 1GHE depicted virtually similar orientation of acetyl-CoA with minor differences in the pose is shown in Figure 8A. Adenine moiety of the acetyl-CoA is present inside the binding pocket of the model, while this moiety is expelled from the cavity

Table 5. ZINC_ID and binding affinities of the best ten compounds with monomer as well as dimer obtained through virtual screening (out of 760 which are obtained from ZINC database through substrate structure based search) (Appendices II-1 and II-2).

Compounds binding at interface of dimer	Binding energy (kcal/mol)	Compounds binding at active site of monomer	Binding energy (kcal/mol)
ZINC65731330	-12.10	ZINC43411693	-10.20
ZINC08215434	-12.10	ZINC85629877	-10.10
ZINC85564657	-12.00	ZINC94303217	-09.90
ZINC85629877	-11.90	ZINC43411702	-09.90
ZINC85564663	-11.80	ZINC08551088	-09.90
ZINC08551088	-11.70	ZINC85574911	-09.80
ZINC85534368	-11.60	ZINC44559643	-09.80
ZINC94303217	-11.60	ZINC43411699	-09.80
ZINC87515504	-11.50	ZINC13829348	-09.80
ZINC85564649	-11.50	ZINC13829346	-09.80

due to the contraction of cavity size by larger size loop in 1GHE. The main-chain contacts include the conserved motif as found in other family members (see residues highlighted in **Figure 2** and labelled residues in **Figure 8C**) and in the current docked complex. The residues, making the direct contact with the substrate acetyl-CoA, are given in **Figure 8B** and the analysis of binding site designates it to be majorly lined by electropositive and hydrophobic residues. In terms of binding energy, screened compounds were found to have high binding affinity even better than substrate acetyl-CoA (**Figure 8D**).

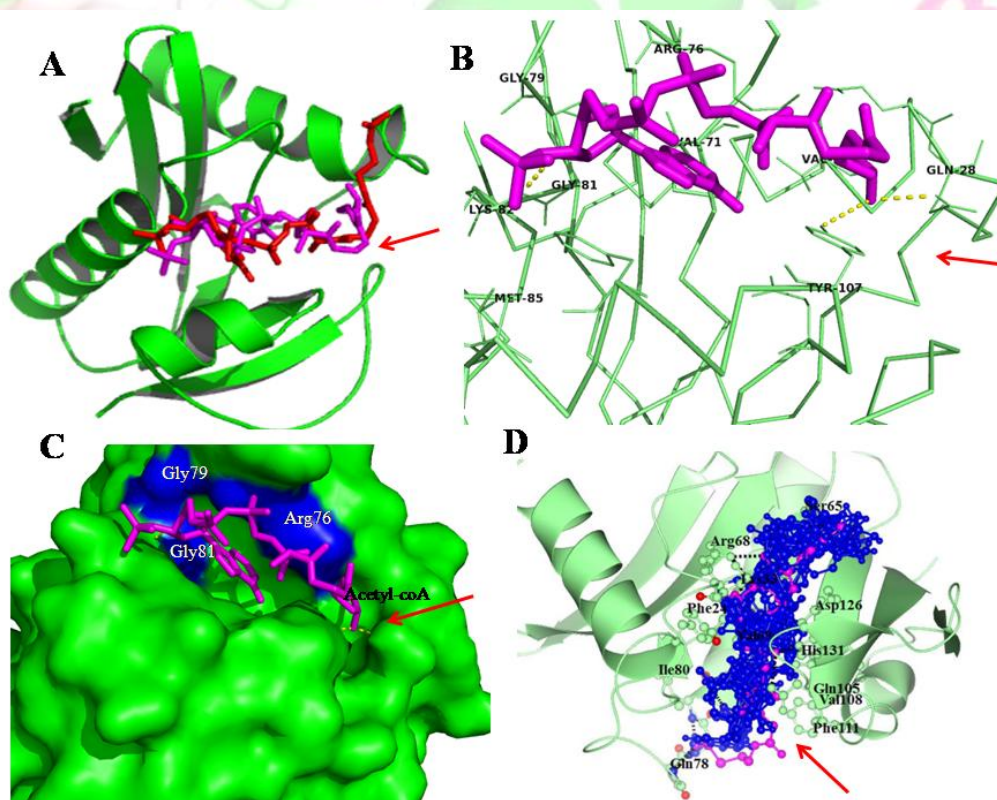


Figure 8 The predicted docking conformation of substrate acetyl-CoA, (docked magenta, stick representation, experimental red). (A) Model (green) showing interactions with acetyl-CoA inside the binding pocket. (B) Ribbon representation of binding pocket residues with substrate. (C) Surface representation of binding pocket with substrate showing GNAT Motif residues (blue). (D) The docking poses of 10 best compounds (blue) superimposed on substrate (magenta).

Due to the similarity in their backbone structure, these compounds could work as competitive inhibitors. Furthermore, in the case of dimeric model, the possible binding modes of substrate as well as substrate-based inhibitors were found to be different, with preference in

binding at interface instead of binding pocket as depicted in **Figure 9B**. The nature of the residues lining the interface is similar to the residues at the active site residues. Binding pose of acetyl-CoA and the inhibitors at the interface are shown in **Figure 9A**.

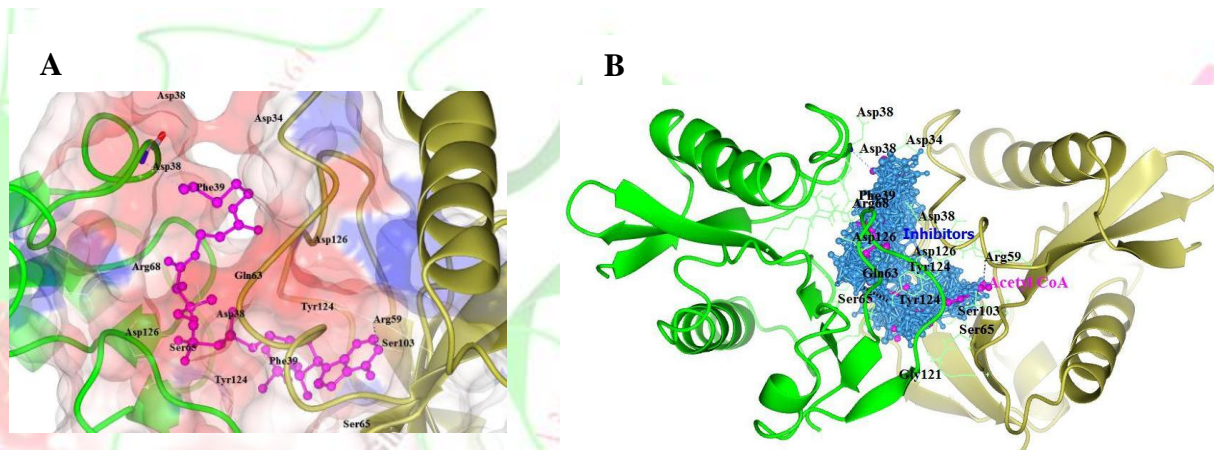


Figure 9. The predicted docking conformations (A) Stick representation of substrate at the dimer interface along with the residue making direct contact in binding. (B) The docking poses of 10 best compounds (blue) superimposed on docked pose of substrate) at the interface (ball and stick representation).

The compounds ZINC65731330, ZINC08215434, ZINC85564657 and ZINC87515504 showed preferential binding at the dimer interface (by 3 kcal/mol) and might be suitable for designing of allosteric inhibitors. Binding of three polyamines whose binding parameters are known in literature [55], was also performed and the result shows that spermine has binding preference over spermidine and putrescine, with putrescine having minimum binding affinity. The docking analysis of aminoglycoside antibiotics with Hpa2 shows presence of significant binding affinity.

Section-II

Acinetobacter baumannii is documented as an opportunistic, multidrug-resistant pathogen. It causes various diseases such as meningitis, pneumonia, cellulitis, bacteremia, endocarditis, osteomyelitis, corneal perforation and urinary tract infections [78–80]. Treatment protocols that are currently employed against *A. baumannii* infections appear to be inadequate, while resistance rates to present therapeutic options such as carbapenems, polymyxins are continuously rising. Therapeutic failure in multidrug resistant pathogens in the current scenario demands the detection of novel drug targets essential at the key steps of metabolic pathways to cure the infections caused by *A. baumannii*. Availability of whole genome sequence of *A. baumannii* [81] facilitates the identification of new non-conventional potential drug targets [82]. Here we have selected the 3-methyladenine DNA glycosylase (TAG) enzyme, a base excision repair glycosylase that recognizes and excises 3-methyladenine (3mA) from DNA, thereby maintaining the genome integrity [83]. DNA bases are constantly damaged and modified by a variety of cellular mechanisms which obstruct vital processes such as DNA replication and transcription. To circumvent this problem, nature evolved many repair pathways in all organisms. For example, base excision repair of modified DNA base is performed by DNA glycosylases by the hydrolysis of *N*-glycosidic bond, resulting in the release of damaged base and leaving the sugar phosphate backbone intact [84]. In prokaryotes, removal of alkylated purine bases through C–N bond cleavage is performed by two DNA glycosylases which are coded by two different genes viz., *tag* and *alkA* genes [85, 86]. The enzymes namely TAG and 3-methyladenine DNA glycosylase II (AlkA), possess capacity to recognize subtle modification in base structure and remove them from the genome. TAG does not show evident primary sequence similarity with other DNA glycosylases due to which its phylogeny is still mysterious. TAG is distinctive from AlkA in substrate specificity and its specificity preserves systematic and constitutive elimination of positively charged 3mA bases [87, 88]. Involvement of TAG enzyme in DNA repair pathway makes it crucial component for the survival of *A. baumannii*. The interesting fact about enzyme TAG is that it has no homologous sequences in archaea, lower eukaryotes, yeast, and in mammals.

Alkylpurine DNA glycosylases have been shown to be essential for the survival of both eukaryotic and prokaryotic organisms. Due to its absence in humans, TAG can be considered as a potential target against *A. baumannii*. Lack of experimental structure for TAG

enzyme motivated us to perform *in silico* and experimental study to provide a structural scaffold theoretically. In this work, we incorporated RepairPDB module of algorithm FoldX with traditional homology modeling procedure to improve the structural quality of model [89]. The need of this extra step during homology modeling is addressed in the methodology and the *Results and discussion section* (2.3). Model energy is evaluated using the Stability module, which includes terms that have been found to be crucial in the definition of protein stability. Subsequently, to gain insight into the residues involved in the identification of a DNA lesion, their excision and in binding specificity, we performed the point mutations using Protein Design utility of FoldX. This demonstrates the specificity of respective interactions at molecular level with reference to the available structural data. The calculated free energy changes show a good correlation with the experimentally observed rate of substrate 3mA excision. Further, structure based virtual screening was carried out to identify possible inhibitors. Theoretical findings were supplemented with the experimental validation of secondary structure and substrate binding studies. Furthermore, we summarize the prime structural and mechanistic characteristics of TAG with a special focus on the residues that play central role in its action. Finally, structure-based inhibitor virtual screening is performed to identify novel anti *A. baumannii* lead compounds.

2.4. Materials and Methods

2.4.1. Computational methods

All computational analyses were carried out on Intel dual core based Microsoft windows XP professional workstation and 16 GB RAM. The workflow for constructing TAG model is depicted in Chart 2.

2.4.2. Homology Modeling of TAG

The reliability of homology modeling depends on the structural similarities between target and the template selected. Thus, identification of correct homologue is decisive for the generation of accurate model. BLASTp was used to search homologous sequences of TAG protein from Protein Data Bank (<http://www.rcsb.org>) and the sequence with highest similarity was used as a template for the homology modeling [90]. Secondary structure prediction and sequence alignment were carried out using ESript [91]. Homology modeling of TAG was performed

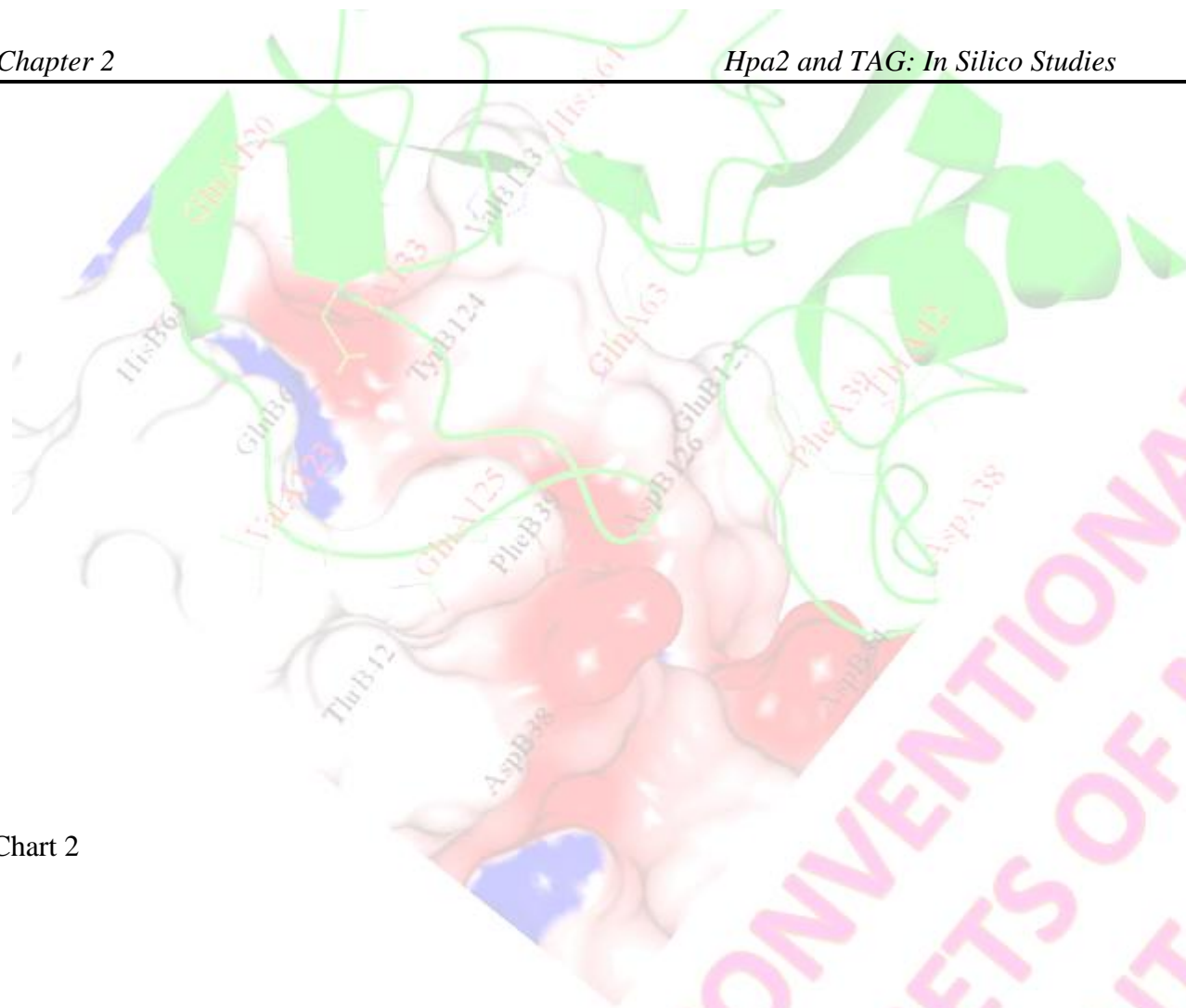


Chart 2

UNCONVENTIONAL PROTEIN
TARGETS OF MULTI DRUG
RESISTANT ACINETOBACTER
JYOTI S. TOMAR
IITR

Using one of the widely accepted 3D structure prediction programs, Modeller9v12 [22]. Modeller constructs the model using a set of restraints derived from the query-template alignment and then energy minimization was done to decrease the violations of these restraints. Restraint-based modeling methodology is better and most flexible as different types of restraints and constraints can be easily incorporated in it. Restraints can be derived from template structure or from experiments like NMR and fluorescence. [92–97]. Initially, 50 models were generated and the model having minimum discrete optimized potential energy (DOPE) score and better structural parameters was chosen for the loop refinement using Chimera Interface.

2.4.3. Model optimization using RepairPDB and evaluation

In general, the first step in model evaluation is examination of structural folds. The accuracy of a fold is ensured by selecting the template optimally aligned with the query sequence. VERIFY3D [24] is used to check the fitness of structure to the sequence [64]. Second basic characteristic of a good model is to have good stereochemistry, for that PROCHECK [27] and WHATIF [25] were used. These programs are widely used for the examination of proper protein stereochemistry, such as symmetry and geometry checks (chirality, bond lengths, bond angles, torsion angles, *etc.*) and structural packing quality. The best model was selected and energy minimization was carried out using Chimera tool Minimize Structure. This was achieved with 1000 steps of the steepest descent algorithm and conjugate gradient algorithm each, using step sizes 0.01 and 0.02, respectively. Minimized model is evaluated for its structural quality using the above mentioned programs.

Comparative study performed by Wallner and Elofsson on the widely acknowledged six modeling programs concludes that Modeller performance is excellent in all aspects while it is inferior in terms of side chain quality [98]. Keeping the problem of the side chain quality in the mind, we used the RepairPDB module of FoldX to further optimize the model. Working style of the command RepairPDB explains how it improves the quality of the model. Initially, it identifies the residues having nonstandard torsion angles and exhibits van der Waals clashes or total energy. Then it mutates the selected residues to alanine and annotates side chain energies of the neighbouring residues. Again it mutates alanine to the original amino acid and recalculates the side chain energies of the same neighbouring residues. Residues having energy difference are explored for different rotamer combination to find new energy minima.

Finally, all side chains are slightly moved to get rid of small steric clashes. In this way RepairPDB quickly removes the small local clashes, and saves computing time by decreasing the number of real rotamer searches. Repaired model is evaluated for its quality using the above mentioned programs. Stability analysis of model and template proteins was performed using energy function of FoldX to compare and analyse the significance of the energy terms contributing to the stability of these proteins. Gibbs free energy of unfolding (ΔG , in kcal mol⁻¹) is calculated using the following equation:

$$\Delta G = \Delta G_{\text{vdW}} + \Delta G_{\text{solvH}} + \Delta G_{\text{solvP}} + \Delta G_{\text{wb}} + \Delta G_{\text{Hbond}} + \Delta G_{\text{el}} + \Delta G_{\text{kon}} + T\Delta S_{\text{sc}} + T\Delta S_{\text{mc}} + T\Delta S_{\text{tr}}$$

where ΔG_{vdW} is summation of van der Waals contributions of all atom, ΔG_{solvH} & ΔG_{solvP} are the differences in solvation energy when going from unfolded to folded state, for non-polar and polar groups, respectively; ΔG_{Hbond} is the free energy difference of intra-molecular hydrogen bond to inter-molecular hydrogen bond, formation (with solvent); $T\Delta S_{\text{mc}}$ & $T\Delta S_{\text{sc}}$ are entropy cost of fixing, the backbone in the folded state and side chain in a particular conformation [68]. The final model is superimposed with the template structure using pymol for calculating the backbone rmsd.

2.4.4. Molecular Docking of substrate and virtual screening of Zinc database in TAG

The ability of molecular docking to predict the structural configuration with optimized and favorable interactions between small molecule and the macromolecular structure is a valuable tool in drug design. Molecular docking was performed using Autodock4.0.2 [27] to reproduce the experimental homologous TAG-substrate/inhibitor complexes. This was undertaken to validate the performance of the software. High-throughput virtual screening (HTVS) programs such as PyRx are useful adjunct to the time consuming and expensive wet bench experiments necessary to discover new drug candidate. AutoDock Vina in PyRx 0.8 (Virtual Screening Tools) is used, which takes receptor as a rigid structure to perform virtual screening. For this study, we collected 550 compounds from ZINC database (<http://zinc.docking.org>) based on at least 60% structure similarity with the substrate 3mA. The energy minimization of the compounds was carried out using 1000 steps of steepest descent prior to virtual screening. Screening parameters include Lamarckian genetic algorithm and grid center at x, y, z coordinates 23.43, 37.62, 35.42 with a grid point spacing of 0.381 Å. After the screening, the top ten compounds showing highest binding affinity are analyzed to get more insight into the stabilizing interactions.

2.5. Results and Discussion

2.5.1. Sequence alignment and secondary structure prediction of TAG

BLAST against the PDB database illustrated that TAG shares at least 53% sequence identity with TAG of *Salmonella typhi* (PDB: 2OFK) and *E. coli* (PDB: 1LMZ) (Table 6). TAG of *A. baumannii* is 186 amino acid long, shows 56% primary sequence identity and 65% similarity with 183 amino acid long TAG of *Salmonella typhi* (PDB: 2OFK resolution 1.50 Å). Therefore, 2OFK is considered as a suitable template for homology modeling (Figure 10). Amino acid sequence of the model indicates that it belongs to 3-methyladenine DNA glycosylase I enzyme, which is also supported by the conserved domain search. Since TAG belongs to superfamily helix–hairpin–helix (HhH) of DNA glycosylases, the HhH motif of TAG is also used for DNA binding in a sequence-independent manner like other protein members of this superfamily [99]. Figure 10 represents sequence alignment of the target and template TAG using ESPript. Inspection of sequence alignment reveals higher sequence similarity in the N-terminal region. The multiple sequence alignment result also demonstrates that residues near N-termini are relatively conserved (Figure 11). It is also known from the literature that functional cavity is mainly mapped by the aromatic residues present near the N-terminal region. Among all the secondary structures [100] the residues present at the helices $\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 4$ and the loops formed from helices $\alpha 1/2$, $\alpha 2/3$, $\alpha 3/4$ and $\alpha 9/10$ are more conserved.

Table 6. Templates from BLAST against PDB with TAG of *A. baumannii*.

PDB id.	Organism	Title	Sequence identity/ Similarity (%)	E-value
2OFK	<i>S. typhi</i>	Chain A, crystal structure of 3-methyladenine DNA glycosylase I (TAG)	56/65	9e ₋₆₆
1LMZ	<i>E. coli</i>	Chain A, solution structure of 3-methyladenine DNA glycosylase I (TAG)	53/65	1e ₋₆₃
2OFI	<i>S. typhi</i>	Chain A, crystal structure of 3-methyladenine DNA glycosylase I (TAG) Bound To DNA 3mA	54/64	9e ₋₆₃
2JG6	<i>S. Aureus</i>	Chain A, crystal structure of a 3-methyladenine DNA glycosylase I from <i>Staphylococcus aureus</i>	35/58	1e ₋₃₉
4AIA	<i>S. Aureus</i>	Chain A, the structural basis of 3-methyladenine recognition by 3-methyladenine DNA glycosylase I (TAG) from <i>Staphylococcus aureus</i>	35/58	1e ₋₃₉

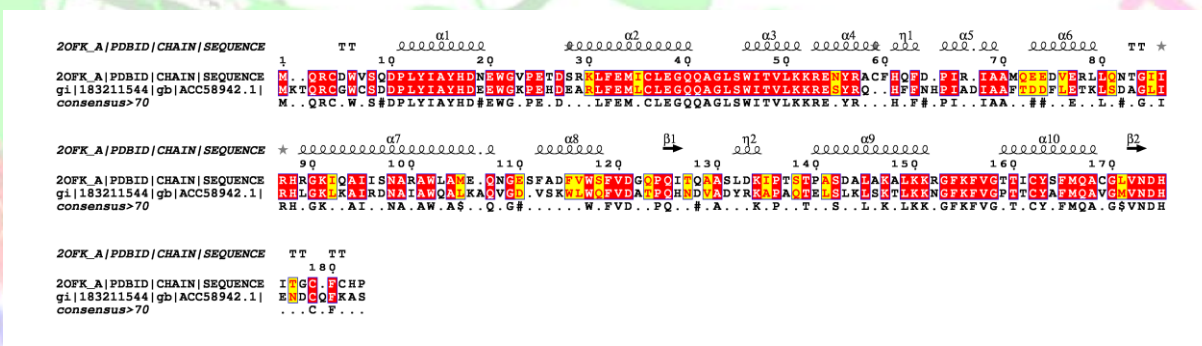


Figure 10. Sequence alignment and secondary structure prediction of the enzymes; TAG (Accession no. YP_001848289.1) and PDB id: 2OFK_A.

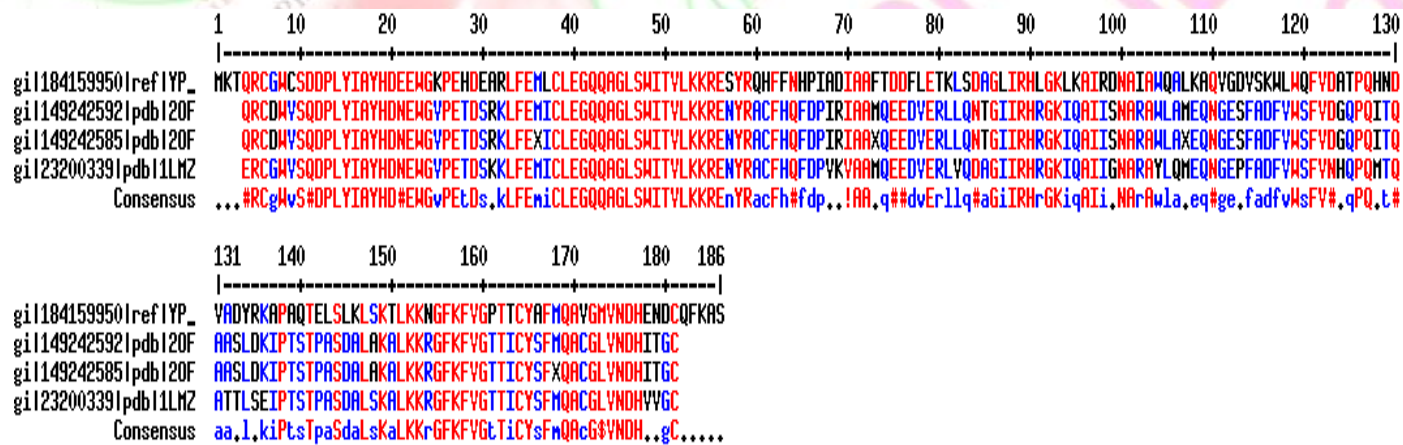


Figure 11. Multiple sequence alignment of query sequence with the template showing sequence similarity (at least 30%).

The loop residues, present between helices $\alpha 5/6$ are not well conserved, but similar nature of these residues in HhH family member might be suggestive of similar functional role. In the TAG from *S. typhi* as well as in *A. baumannii*, the residues Cys181 and Cys4 are conserved and both are present side by side in space to make zinc binding motif. However, this conserved Cys residue is not found in its closely related bacterial DNA glycosylase AlkA and Helicobacter pylori 3mA DNA glycosylase (MagIII) enzymes. The C-terminal residues contribute in DNA holding during repair process through hydrogen bond interactions to the backbone phosphate and participate in the formation of zinc snap motif.

2.5.2. Validation of model and its optimization using FoldX RepairPDB

Verification of model quality and estimation of the probability and magnitude of errors are critical in advancing the state of homology modeling [101–103]. As discussed in the *Methodology section* (2.4), for optimization, the model has been passed through repair function of FoldX algorithm. The repair function can be considered as one of most successful combinatorial optimization algorithms in generating rotamer ensemble having favourable conformations. Assessments of the stereochemical parameters and 3D-1D score for fitness of model structure to query sequence illustrates that quality of model generated by the Modeller9v12 is not optimal (Table 7). Similarly, structure validation reports of the energy minimized model were also found disagreeable (Table 7), which may be due to the drifting away of the coordinates from the template coordinate during energy minimization. The quality of FoldX repaired model is also checked using above mentioned validation programs and interestingly the model quality was found to be optimal and comparable to template structural parameters (Table 7). This demonstrates that repair function is successful in optimizing model parameters. Further, it saves computational time by decreasing the number of real rotamer searches. Ramachandran plot of the optimized model structure shows that more than 91.1% of the residues have ϕ and ψ angles in the core regions, 8.9% residues in the allowed regions, with no residues in the generously allowed and disallowed regions. The percentage of residues having averaged 3D-1D score > 0.2 was improved from 74 to 88 by the repair function revealing successful fitting of rotamers performed by the same. The rmsd value for the model is calculated and found to be 0.19 Å by keeping the template structure 2OFK as reference. These results suggest that quality of the model (id: lldbpmi6whn4df15v; Model Archive doi: ma-a6bjp) is good and sufficiently accurate for further molecular modeling studies.

Table 7. Summary of structural validation programs and parameters values for models and template (pdb id. 2OFK_A). Initial model (model obtained from MODELLER 9v12), minimized model (model after minimization) and repaired model (model obtained after FoldX repair function).

Server	Parameters	Initial model	Minimized model	FoldX repaired model	Template
PROCHECK	Residues in core, Allowed	91.4%, 8.6%	88.1%, 10.7%	91.1%, 8.9%	92.6%, 7.4%
	Generously allowed regions	0.0%	0.6%	0.0%	0.0%
	Labelled residues	4	7	4	3
PROSA	PROSA Z-score	-6.90	-6.79	-6.98	-7.79
VERIFY3D	Residues having averaged 3D-1D score > 0.2	74.33%	84.43%	88.24%	95.69%
ERRAT	Overall quality factor	85.39%	91.70%	91.87%	92.39%
PDBeFold	Q score	0.94	0.94	0.94	
	RMSD	0.30 Å	0.68 Å	0.19 Å	

The FoldX stability analysis upon execution produces Gibbs energy of folding decomposed in various energy terms; it is calculated as difference in free energy between folded and unfolded states. The values for energy terms ΔG_{vdW} , ΔG_{solv} , ΔG_{solvP} and ΔG_{Hbond} , of each atom type, have been derived from experimental data while the entropic costs (ΔS_{mc} and ΔS_{sc}) are derived from theoretical estimation. Values of different energy terms calculated for the repaired TAG model and template pdb structure are given in supplementary information (Table 8). Contributions coming from destabilizing energy terms in the model are found to be balanced by favourable energy terms. Molecular interactions contributing to the favourable energy are dominated by H-bonds, hydrophobic and van der Waals, while entropic loss and van der Waals clashes are found to be main destabilizing interactions in the model. Stability

Table 8. Comparison of various energy terms obtained from Stability analysis command of the FoldX, for the model with its experimental homologous structures (template).

PDB/Model	ΔG_{total}	ΔG_{H-bond}	ΔG_{vdw}	$\Delta G_{solvH+solvP}$	Entropy sum	ΔG_{vdw} clashes
OFK	-40.15	-188.4	-222.0	+30.0	+381.4	+6.3
2OFI	-32.4	-181.3	-221.7	+34.4	+367.16	+9.5
Model	-28.6	-192.8	-223.6	+18.7	+386.7	+10.23

analysis of template *S. typhi* TAG is also performed, to comparatively judge the model quality. Interestingly the energy terms of the repaired model are near to template 2OFK.

2.5.3. Comparative structural analysis of model with TAG of *S. typhi*

The current model consists of ten α -helices, two 3_{10} -helices and one parallel β -sheet, sequence of secondary structure in the model is $\alpha 1\alpha 2\alpha 3\alpha 4\eta 1\alpha 5\alpha 6\alpha 7\alpha 8\beta 1\eta 2\alpha 9\alpha 10\beta 2$. The secondary structures $\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 4$ and loops formed from helices $\alpha 1/2$, $\alpha 2/3$, $\alpha 3/4$ and $\alpha 9/10$ are more conserved. These secondary structures play an important role in DNA binding and base excision mechanism and this conservation in secondary structure elements elucidate the closer structural-functional resemblance of model with template (Figure 10). The analysis of secondary structure shows that the enzyme consists of 62.4% helix, 3.2% sheet, 8.6% turn and 25.8% coil structure. Residues making plug and wedge [104] are present on the same secondary structure element ($\alpha 2/3$ loop) of the model (contrary to the observation in other HhH glycosylases). Residues Gly43–Leu44 from TAG of *S. typhi* involved in conserved intercalation mechanism are significant for the interrogation of lesion in DNA during search process. These residues make special interactions leading to a 60° – 70° bend in the DNA, helping the TAG enzyme in gaining access to the modified base without any significant conformational changes. These residues are known to play pivotal role in holding DNA during enzyme catalysis [105] and are also present at $\alpha 2/3$ loop (Gly45–Leu46) of the TAG model. The Zn^{2+} binding motif is conserved and it tethers the N- and C-termini. Cao and co-workers proposed that the zinc snap motif of TAG contributes for the stabilization of HhH domain in a way similar to that of protein–protein interaction present in larger HhH superfamily members [106]. In the model, 3mA binding pocket is located at the interface between the Zn^{2+} binding domain and helical domain with HhH motif. In TAG, HhH motif is known to contribute ~50% of polar interactions made with DNA backbone and the residues involved in such interactions (except Ile161) are present at the same secondary structure in the model as in template TAG enzyme (Figure 10). The electron-rich aromatic binding pocket of TAG drives a significant amount of binding energy through π – π stacking interactions with electron-deficient alkylated base (3mA) [86]. The electronegative pocket in the TAG provides the compatible environment to positively charged 3mA (Appendix II-3). The 3mA found to form two H-bonds with Glu38 residue and one H-bond with Tyr16 in experimental *S. typhi* TAG. These residues work as adenine recognition motif, homologous to Glu37 and Gln182 residues of the MutY adenine

glycosylase enzyme. High quality function score (Q score = 0.94) from PDB-efold, one of the well known structure alignment programs, affirms resemblance between the fold of the aligned molecules (Table 7). Superposition of *S. typhi* TAG and model structures shows that the protein backbones and positions of active site residues are highly superimposing (with an rmsd of 0.16 Å for all main-chain atoms and 0.10 Å for active site atoms (Figure 12).

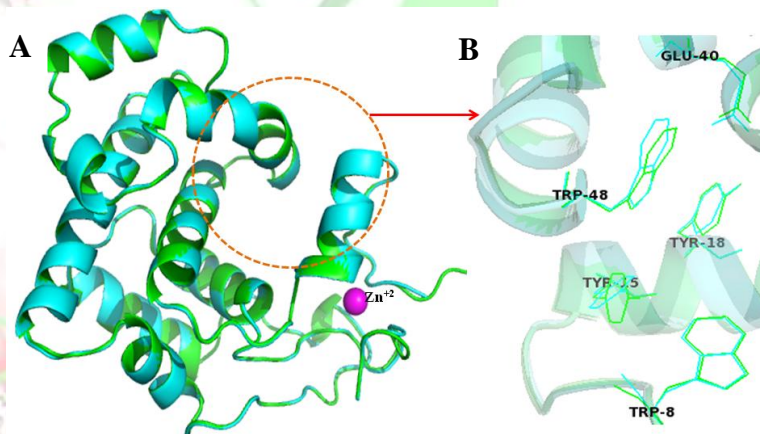


Figure 12. Structural alignment of the enzymes; predicted *in silico* TAG model (coloured green) and template (PDB id: 2OFK_A) (coloured cyan) (A) Complete TAG model structure. (B) Residues lining the model binding pocket superimposed on experimental residues of binding pocket.

2.5.4. Molecular docking analysis of substrate and virtual screening to identify inhibitors against TAG enzyme

As the very first step, molecular docking of co-crystallized ligands was performed in the macromolecular structures. The purpose of this redocking was to examine whether molecular docking simulation using Autodock4 are able to reproduce the binding pose as reported in experimental TAG protein complexes and to characterize the molecular recognition interactions. Homologous TAG protein complexes present in the PDB were also used in redocking procedure to understand the molecular recognition interactions. The rmsd values are calculated by keeping the experimental pose as reference (Appendix II-4).

In redocking, the atomic structures from a co-crystal are used as starting coordinates. The rmsd value for 3mA is 1.67 Å (Figure 13A) and for other experimental complexes are approximately 2 Å. This signifies that the autodock is successful in reproducing experimental pose. Theoretical binding energies, rmsd values and IC₅₀ values for all the redocked ligands are given in Table 9. Theoretical binding energy is showing good correlation with the

experimental IC₅₀ values, which indicates that the autodock is successful in discriminating various TAG enzyme complexes. Further, analysis of binding sites, done for experimental complex 2OFI to examine the molecular interactions, indicates that the residues Tyr16 and Tyr13 are involved in the direct hydrogen bonding and play crucial role in tuning of redocked pose (Figure 13A). Redocked pose of 3mA also forms three H-bonds – two with Tyr13 and one with Tyr16 side chain (Figure 13A, Blue coloured). Other contribution for binding is coming from van der Waals interactions and π - π stacking with a conserved Trp side chain. Molecular docking analysis of the substrate 3mA with the model was also performed which shows that docked pose is slightly different from the experimental pose. Similar to crystalized conformation in 2OFI the docked pose of 3mA forms H-bonds with the residues Tyr15 and Tyr18 (Figure 13B and Figure 13C). The possible reason for this difference in the docked pose of 3mA in template as well as in model from experimental conformation might be due to the lack of conformational flexibility of receptor during molecular docking simulation. Experimental results (our CD data as well as literature data) show that binding and removal of

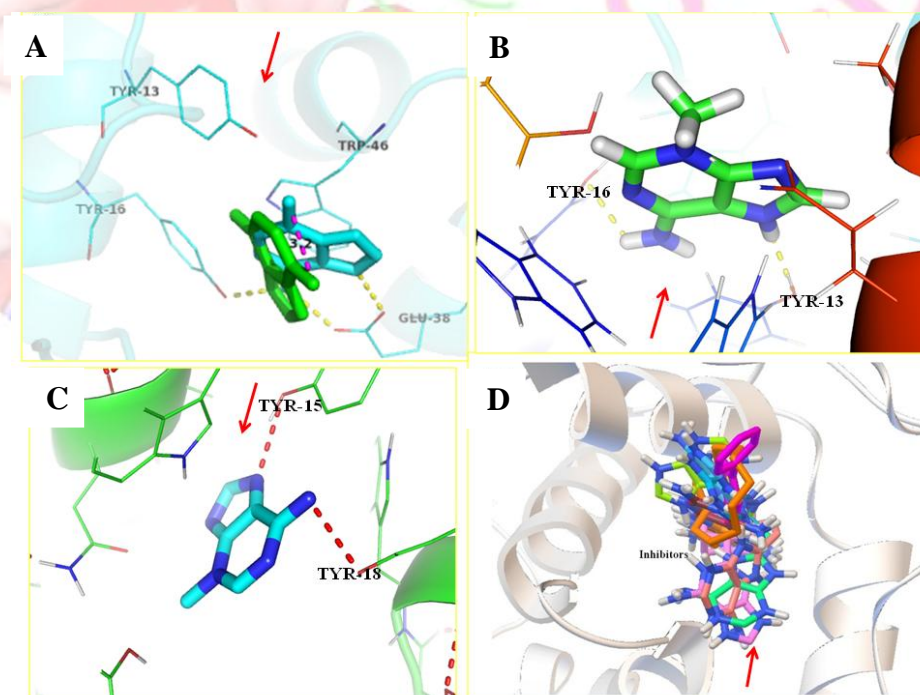


Figure 13. (A) Active site showing residues forming H-bonds with experimental (cyan) and redocked (green) poses of substrate 3mA. (B) Template active site showing residues crafting H-bonds with redocked substrate 3mA. (C) Model active site residues involved in the H-bonding with docked 3mA. (D) Model active site showing best inhibitors and docked 3mA.

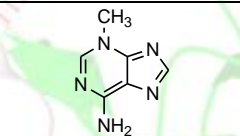
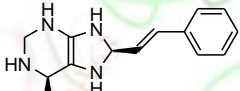
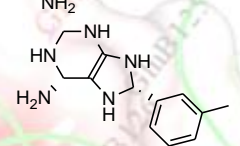
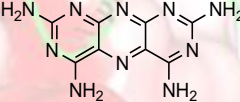
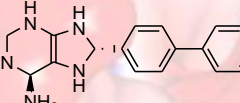
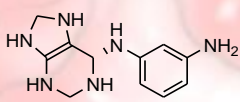
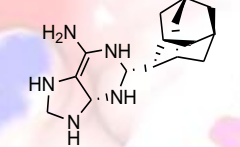
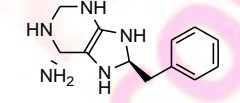
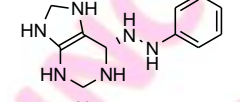
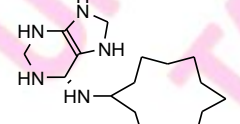
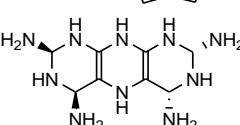
Table 9. Binding energies (kcal/mol), rmsd values and binding interactions (Active site residues involved in the H-bond, Hydrophobic interactions) of experimental TAG complexes determined using Autodock 4.2.

PDB id	Binding energy	IC50 (experimental)	RMSD (Å)	H-bond	Hydrophobic Int.
2OFI	-4.98	1.5 ± 0.3 mM	1.67	Y16, E38	W6, W46, S164
1PVS	-6.06		1.52	Y239, L25	W218
1PU7	-5.16		0.66	-	W24, P45, L211
1PU8	-6.98	50 ± 10 µM	1.09	L211	W24, P26
1P7M	-5.18	1.5 ± 0.3 mM	2.12	Y16, E38	W46, W21, Y13
1AIA	-4.76		2.06	Y16, E38	W46, P6
1AI5	-4.46		1.96	E38	W46, A168

3mA is completed by the conformational change in the enzyme, which may help in optimization of substrate binding with enzyme (Figure 13A, Green coloured).

The virtual screening has been successfully used as an adjuvant to high throughput screening approaches for the discovery and development of new compounds. In the present study, the information obtained about the molecular interactions between substrate and enzyme required for the optimal binding is used in the virtual screening. Here, all the 550 compounds selected for virtual screening exhibited binding with target protein. The main objective of this part of work is to identify potent inhibitors of TAG enzyme using a combination of various *in silico* techniques like redocking of experimental complexes, virtual screening and ADMET. The top ten compounds based on binding affinity are summarized in Table 10, and these compounds are further analysed to inspect the active site residues and molecular interactions involved in binding to enzyme. Stabilizing interactions for the inhibitors are provided by the residues Trp8, Tyr18, Glu40, Gln43, Trp48 and Gln169, which are known to be crucial for catalytic activity using the experimental mutational study [105]. H-bonds, van der Waals and stacking interactions play crucial role in the binding of 3mA and inhibitors. Finally, bound conformations of inhibitors having high binding affinity are superimposed to check the deviation in the binding poses (Figure 13D). The evaluation of drug likeness score for these molecules was done using web server <http://www.molinspiration.com/> (Table 10).

Table 10. Binding affinities (kcal/mol), and H-bond forming active site residues of the best ten compounds out of 550 compounds obtained from ZINC database having at least 60% structure similarity with the substrate 3mA. The Drug likeness score is given (higher drug likeness score designates the compound to be active drug).

Inhibitors	Compound structure	Binding affinity	Residues (No. of H-bonds)	Drug likeness score
3mA		-4.98	Y16, E38 (3-bonds)	-
1		-8.01	Y15, S47, W48 (3-bonds)	-0.09
2		-7.90	Y15 (2-bonds)	0.73
3		-7.62	W8, Y15, Q169 (6-bonds)	-1.79
4		-7.61	Y15, E40 (2-bonds)	1.03
5		-7.57	Y15, E40, Q43 (6-bonds)	-0.74
6		-7.55	W8, Y15 (3-bonds)	1.73
7		-7.54	Tyr15 (2-bonds)	1.63
8		-7.52	Y15, Y134 (3-bonds)	0.43
9		-7.51	Y15, Q169 (3-bonds)	-3.72
10		-7.50	W8, Y15, Q169 (6-bonds)	-1.19

It calculates the score as a complex balance between structural features and various molecular properties to determine whether the molecule taken for study is comparable to the well known drugs. The high value of the drug likeness score symbolizes the higher probability of a particular molecule to be active. Based on the score values, the compounds **4**, **6** and **7** qualify to be the better candidates for further experimental studies; binding pose of these inhibitors are similar with the substrate 3mA and makes similar type of molecular interaction with the active site residues. Here we conclude that the compounds **4**, **6** and **7** exhibit drug-like properties, and exhibit significant binding affinity and optimal molecular interactions with the enzyme TAG, and due to which these compounds could be used as competitive inhibitors. However, experimental enzyme inhibition assay is required to reconfirm the inhibitory potential. Synthesis of these compounds is in progress in our laboratory.

2.5.5. Mutational studies of amino acid residues involved in the TAG enzyme activity

To get more insight about amino acids directly or indirectly involved in TAG activity, we performed point mutations using Protein Design and Position scan utilities of FoldX. Generally mutations are performed in the wet lab to study the contribution of an amino acid on protein stability and functionality. Such site-directed mutagenesis in the wet lab is laborious and costly. Metz and co-workers [105] used twelve residues from TAG of *S. typhi* in their experimental study. Similarly, we also selected those twelve residues from enzymes TAG of *S. typhi* and *A. baumannii* for the mutational study. The calculated $\Delta\Delta G$ values for mutation of twelve residues in both the enzymes are given in Table 11 and displayed in Figure 14A.

The calculated change in ΔG values for all the twelve residues from TAG of *S. typhi* (only positive values) is in good agreement with experimental 3mA excision rate ($10^{-3} \text{ min}^{-1} \text{ mM}^{-1}$) (Figure 14B). Similar pattern in the change in free energy values for the TAG of *A. baumannii* is also obtained (Figure 14A). The calculated change in free energy values for the residues Glu40 and Tyr18 is very high ($\sim 3.55 \text{ kcal/mol}$), which indicates decrease in the stability upon mutation of these residues. There are two possible reasons for the loss of activity on mutation of residues Glu40 and Tyr18 (~ 2.05 and 1.5 kcal/mol ; respectively). Firstly, the replacement of polar 'hydrogen bond-forming' hydroxyphenyl group of tyrosine with electron-rich phenyl moiety of phenylalanine may cause deformations in the perfectly shaped active site of TAG which results in positive free energy change and the loss of activity [105] as shown in Figure 14A.

Table 11. Mutational studies of the residues that are important for the enzyme activity, calculation of $\Delta\Delta G_{cal}$ (WT-Mut) values are given for model as well as template (2OFK). Calculated $\Delta\Delta G_{cal}$ (WT-Mut) values are compared with the experimental 3mA excision rate.

Model Residue	$\Delta\Delta G$ (kcal/mol)	Template Residue	$\Delta\Delta G$ (kcal/mol)	3mA Excision rate $10^{-3} \text{ min}^{-1} \mu\text{M}^{-1}$
W8A	-2.61	W6A	-3.02	29.2 ± 16.9
Y15F	+1.08	Y13F	+1.07	54.7 ± 39.3
Y18F	+1.53	Y16F	+1.33	19.0 ± 4.60
E40A	+2.04	E38A	+2.13	0.7 ± 0.1
Q43A	+0.93	Q41A	+0.60	36.2 ± 3.1
G45L	-2.09	G43L	-1.42	2.5 ± 0.8
L46A	-2.86	L44A	-1.03	6.5 ± 4.4
S47A	+0.73	S47A	+0.46	230 ± 7.8
W48A	-2.48	W46A	-4.12	20.7 ± 7.4
K93A	-2.11	K91A	-1.47	26.9 ± 1.5
T162V	+0.34	T160V	+0.28	273.8 ± 68.3
Q169A	+0.64	Q167A	+0.85	107.2 ± 55.7

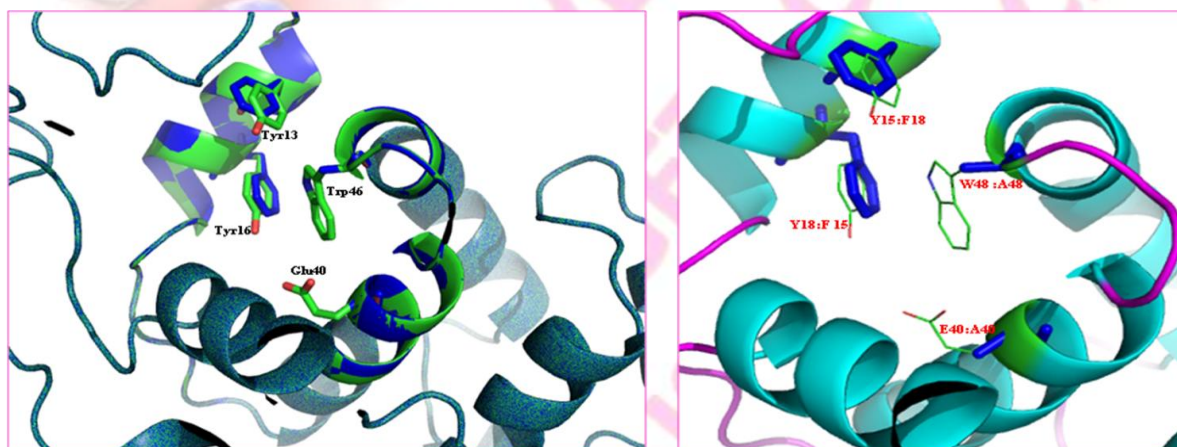


Figure 14A. The active site of template (2OFK) and model showing residues considered for the mutational study, model active site residues before (green) and after (blue) the mutation are displayed as lines and stick, respectively.

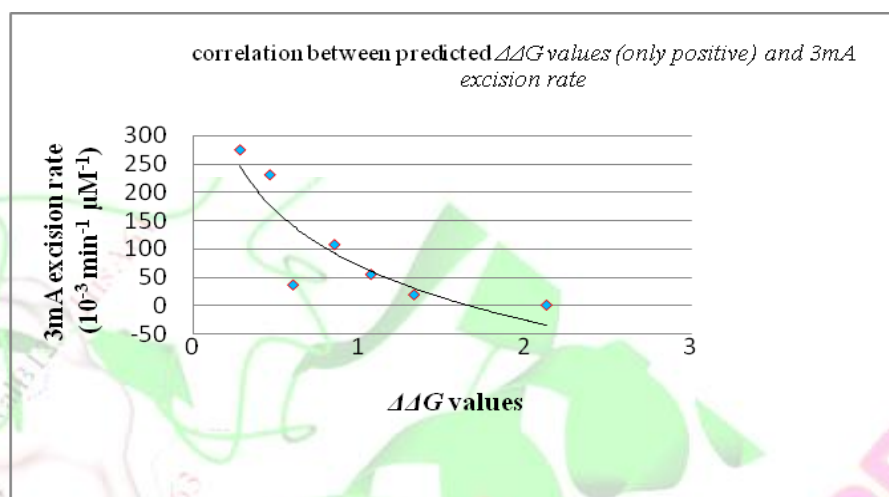


Figure 15B. The graph showing correlation between the theoretical change in $\Delta\Delta G$ value versus experimental 3mA excision rate.

Second reason is the loss of H-bonds present at the Hoogsteen and Watson–Crick faces of the substrate 3mA, as known from our molecular docking analysis and the crystal structure [105, 107–109]. The obtained large positive $\Delta\Delta G$ value (~ 2.05 kcal/mol) for the residue Glu38Ala indicates that glutamic acid plays a major role in providing the binding specificity for 3mA. Similar to Cao and his co-workers [110], our calculated $\Delta\Delta G$ values indicate that mutations of residues Trp6, Tyr13, Tyr16, Glu38, Gln41, Ser47, Thr160 and Gln167 in *S. typhi* TGA and homologous residues Trp8, Tyr15, Tyr18, Glu40, Gln43, Ser47, Thr162 and Gln169 in *A. baumannii* TGA modulate the enzymatic function by destabilizing the pocket geometry and interactions crucial for binding of substrate 3mA.

However, mutation of Trp6, Gly43, Leu44, Trp46 and Lys91 showed negative $\Delta\Delta G$ values, indicating stabilizing effect of respective mutations. In the case of Trp6Ala and Trp46Ala mutations, the value of free energy change is contributed by van der Waals interactions and solvation terms. Mutations of Trp to Ala are energetically favourable because replacement of indole ring through methyl group at the β -position may impose rigidity, means mutation of Trp to Ala produces the rigidity in the active pocket and loss of conformational freedom could be adding to stability. Electron-rich side chains of Trp present at the binding pocket of TAG enzyme make π – π stacking interactions with the electron-deficient nucleobase 3mA, needed for the enzyme activity. On mutation of Trp to Ala, the activity is lost because of the unavailability of the aromatic side chain to make π – π stacking interaction (Figure 14A). The plug and wedge residues Gly43 and Leu44 are important for the DNA conformation

stabilization during accommodation of extra helical nucleotide. The change in free energy values for both of Gly43 and Leu44 residues on mutation is negative indicating that these mutations impart stability in the TAG. In both the TAGs and in its related member AlkA, Leu is present at the respective position which is surrounded by hydrophobic residues. Due to which, substitution of Leu44 side chain (Leu44→Ala44) considerably increases the stability while sacrificing its functionality. The $\Delta\Delta G$ values of mutation of residues are in agreement with literature data. Analysis of molecular interactions indicates how these residues play key role in the activity and how they help in the maintenance of pocket shape and environment suitable for 3mA binding. Effects of mutations are more dramatic for the key catalytic residues and replacement of the residues holding the DNA appends stability pay off.

2.6. Conclusion

Three-dimensional structures of monomer Hpa2 from *A. baumannii* were generated using the modified protocol in the first section (2.2). The results show that irrespective of sequence similarity, Hpa2 flaunts the GNAT fold, characteristic of GNAT superfamily. It shares a common catalytic site with 1Q2Y (*B. subtilis*), 3EFA (*L. plantarum*) and 1XEB (*P. aeruginosa*), revealing significant similarity in the binding pocket and substrate acetylation mechanism in the members of this superfamily. Dimeric model was constructed using experimental information and validated for structural parameters and stability. Moreover, the developed protocol is quite easy to use for the construction of dimeric model using the available experimental information (mutation, sequence similarity, biophysical/biochemical). Analysis of the dimeric interfaces of 1QSM (*S. cerevisiae*), 1XEB and the model using *Analyse complex* of FoldX was performed, which shows that the dimerization of Hpa2 commences in the prokaryotes and tight interactions at interface is acquired during the course of evolution. Furthermore, analysis of binding pose of inhibitors screened by PyRx showed consistency in the pattern of stabilizing forces for all of the compounds; however, inhibition experiments are required to reconfirm this prediction. Binding analysis of Hpa2 with polyamines indicates that spermine has more binding affinity than spermidine, whereas putrescine has the least binding affinity which is in line with the reported biochemical studies. Described herein is application of Hpa2 structure-based screening strategy to identify novel antibacterial compounds.

In the second section (2.4), the three dimensional structure of TAG from *A. baumannii* was constructed. Model structural analysis results show that TAG shares the HhH domain as is characteristic of DNA binding protein. It shares a common catalytic site with the TAG from *Salmonella typhi* and hints that it has some unique features which are different from higher eukaryotes adenine DNA glycosylases and this makes it a potential target for drug design against *A. baumannii* and to overcome the multi-drug resistance. The alignment of TAG sequences indicates that the key residues involved in polar interactions are conserved in query and all templates. Analysis of redocking methodology confirms that binding mode and H-bond pattern of various TAG-complexes are almost similar except for the H-bonds formed with Glu residue. The binding pose of inhibitors screened shows consistent pattern of stabilizing forces as noticed in redocked experimental complexes. The obtained TAG inhibitors (compounds **4**, **6** and **7**) can be used for designing of the more potent and selective binding inhibitors against HhH family and may have capacity to inhibit TAG from other related species also, but enzyme inhibition assays are required to confirm this prediction. The molecular docking analysis revealed that Glu40 and Tyr18 are conserved and catalytically important for the enzyme function. This was further confirmed by the large positive $\Delta\Delta G$ value obtained from the mutagenesis experiments.

- In the performed *in silico* studies, we found that RepairPDB of FoldX command further improves the model generated using the Modeller9v12. Here we found notable increment in the Verify3D score and ERRAT score which indicate that RepairPDB improves the fitness of residues in the 3D structure. It also improves the stereochemical property of the residues by selecting suitable pair of rotamers and local minimization. Further, it helps in the removal of clashes present in the model structure. The foremost advantage of using RepairPDB of FoldX command is that it takes only a few minutes time to complete the run. Due to this fact, one can try it for the model refinement which may lead for the structure improvement; otherwise there will not be any loss of time.
- In this study, FoldX command is used for model refinement as well as for assessment. Stability, Analyse Complex, Protein Design and Position scan modules of FoldX were used to get energetics, interface residues of model and to study the effect of point mutations on the stability and activity of enzyme.

2.7. References

1. Berman, H., Henrick, K., Nakamura, H. and Markley, J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data; *Nucleic Acids Res.* 35, D301–D303.
2. Remmert, M., Biegert, A., Hauser, A. and Soding, J. (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment; *Nat. Methods.* 9, 173–175.
3. Baker, D. and Sali, A. (2001) Protein structure prediction and structural genomics; *Science* 294, 93–96.
4. Das, R. and Baker, D. (2008) Macromolecular modeling with rosetta; *Ann. Rev. Biochem.* 77, 363–382.
5. Ward, A.B., Sali, A. and Wilson, I.A. (2013) Biochemistry. Integrative structural biology; *Science* 339, 913–915.
6. Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. and Jones, D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life; *J. Mol. Biol.* 337, 635–645.
7. Chothia, C. and Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins; *EMBO J.* 5, 823–826.
8. Guex, N., Peitsch, M.C. and Schwede, T. (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective; *Electrophoresis* 30, S162–S173.
9. Peitsch, M.C. (1995) Protein Modeling by E-mail; *Nat. Biotech.* 13, 658–660.
10. Sanchez, R. and Sali, A. (1998) Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome; *Proc. Natl. Acad. Sci. USA* 95, 13597–13602.
11. Sutcliffe, M.J., Haneef, I., Carney, D. and Blundell, T.L. (1987) Knowledge based modelling of homologous proteins, Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures; *Protein Eng.* 1, 377–384.
12. Worth, C.L., Gong, S. and Blundell, T.L. (2009) Structural and functional constraints in the evolution of protein families; *Nat. Rev. Mol. Cell Biol.* 10, 709–720.
13. Kryshtafovych, A. and Fidelis, K. and Moult, J. (2009) CASP8 results in context of previous experiments; *Proteins* 77 (Suppl 9), 217–228.
14. Rost, B. (1999) Twilight zone of protein sequence alignments; *Protein Eng.* 12, 85–94.
15. Sanchez, R. and Sali, A. (1997) *J. Mol. Struct. (Theochem)* 398, 489–496.
16. Holm, L. and Sander, C. (1996) The FSSP database: fold classification based on structure-structure alignment of proteins; *Nucleic Acids Res.* 24, 206–209.
17. Sander, C. and Schneider, R. (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment; *Proteins* 9, 56–68.
18. Holm, L. and Sander, C. (1998) Touring protein fold space with Dali/FSSP; *Nucleic Acids Res.* 26, 316–319.

19. Cuff, A.L., Sillitoe, I., Lewis, T., Redfern, O.C., Garratt, R., Thornton, J. and Orengo, C.A. (2009) The CATH classification revisited-architectures reviewed and new ways to characterize structural divergence in superfamilies; *Nucleic Acids Res.* 37 (Database issue), D310–314.
20. Guex, N., Peitsch, M.C. and Schwede, T. (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective; *Electrophoresis* 30(S1), S162–S17.
21. Levitt, M. (1992) Accurate modeling of protein conformation by automatic segment matching; *J. Mol. Biol.* 226, 507–533.
22. Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints; *J. Mol. Biol.* 234, 779–815.
23. Linge, J.P., Williams, M.A., Spronk, C.A., Bonvin, A.M. and Nilges, M. (2003) Refinement of protein structures in explicit solvent; *Proteins* 50, 496–506.
24. Fan, H. and Mark, A.E. (2004) Refinement of homology-based protein structures by molecular dynamics simulation techniques; *Protein Sci.* 13, 211–220.
25. Nakajima, N., Higo, J. and Kidera, A. (2000) Nakamura, H. Free energy landscapes of peptides by enhanced conformational sampling; *J. Mol. Biol.* 296, 197–216.
26. Lee, M.R., Tsai, J., Baker, D. and Kollman, P.A. (2001) Molecular dynamics in the endgame of protein structure prediction; *J. Mol. Biol.* 313, 417–430.
27. Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR; *J. Biomol. NMR.* 8, 477–486.
28. Hennig, J., Wang, I., Sonntag, M., Gabel, F. and Sattler, M. (2013) Combining NMR and small angle X-ray and neutron scattering in the structural analysis of a ternary protein-RNA complex; *J. Biomol. NMR.* 56, 17–30.
29. Kalinin, S., Peulen, T., Sindbert, S., Rothwell, P.J., Berger, S., Restle, T., Goody, R.S., Gohlke, H. and Seidel, C.A. (2012) A toolkit and benchmark study for FRET-restrained high-precision structural modelling; *Nat. Methods* 9, 1218–1225.
30. Ochi, T., Wu, Q., Chirgadze, D.Y., Grossmann, J.G., Bolanos-Garcia, V.M. and Blundell, T.L. (2012) Structural insights into the role of domain flexibility in human DNA ligase IV; *Structure* 20, 1212–1222.
31. Costanzi, S. (2013) Modeling G protein-coupled receptors and their interactions with ligands; *Curr. Opin. Struct. Biol.* 23, 185–190.
32. Oshiro, C., Bradley, E.K., Eksterowicz, J., Evensen, E., Lamb, M.L., Lanctot, J.K., Putta, S., Stanton, R. and Grootenhuys, P.D. (2004) Performance of 3D-database molecular docking studies into homology models; *J. Med. Chem.* 47, 764–767.
33. McGovern, S.L. and Shoichet, B.K. (2003) Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes; *J. Med. Chem.* 46, 2895–2907.
34. Kairys, V., Fernandes, M.X. and Gilson, M.K. (2006) Screening drug-like compounds by docking to homology models: a systematic study; *J. Chem. Inf. Model.* 46, 365–379.

35. Damm-Ganamet, K.L., Smith, R.D., Dunbar, J.B., Jr, Stuckey, J.A. and Carlson, H.A. (2013) CSAR Benchmark Exercise 2011–2012: Evaluation of Results from Docking and Relative Ranking of Blinded Congeneric Series; *J. Chem. Inf. Model.* 53, 1853–1870.
36. Kufareva, I., Rueda, M., Katritch, V., Stevens, R.C. and Abagyan, R. (2011) Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment; *Structure* 19, 1108–1126.
37. SAMPL. Proceedings of the Third Annual Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) Challenge and Workshop. June 2009. Montreal, Canada; *J. Comput. Aided Mol. Des.* (2010) 24, 257–383.
38. Ripphausen, P., Nisius, B., Peltason, L. and Bajorath, J. (2010) Quo vadis, virtual screening? A comprehensive survey of prospective applications; *J. Med. Chem.* 53, 8461–8467.
39. Vries, S.J.D., Dijk, M.V. and Bonvin, A.M.J.J. (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* 5, 883–897.
40. Box, G.E.P., and Draper, N.R. (1987) Empirical model-building and response surfaces. New York: John Wiley.
41. Schwede, T., Sali, A., Honig, B., Levitt, M., Berman, H.M., Jones, D., Brenner, S.E., Burley, S.K., Das, R. and Dokholyan, N.V., Dunbrack, R.L., Fidelis, K., Fiser, A., Godzik, A., Huang, Y.J., Humblet, C., Jacobson, M.P., Joachimiak, A., Krystek, S.R., Kortemme, T., Kryshtafovych, A., Montelione, G.T., Moulton, J., Murray, D., Sanchez, R., Sosnick, T.R., Standley, D.M., Stouch, T., Vajda, S., Vasquez, M., Westbrook, J.D. and Wilson, I.A. (2009) Outcome of a workshop on applications of protein models in biomedical research; *Structure* 17, 151–159.
42. Haas, J., Roth, S., Arnold, K., Kiefer, F., Schmidt, T., Bordoli, L. and Schwede, T. (2013) The Protein Model Portal - a comprehensive resource for protein structure and model information; *Database (Oxford)*, bat031.
43. Joly-Guillou, M.L. (2005) Clinical impact and pathogenicity of *Acinetobacter*; *Clin. Microbiol. Infect.* 11, 868–873.
44. Murray, C.K. and Hospenthal, D.R. (2005) Treatment of multidrug resistant *Acinetobacter*. *Curr. Opin. Infect. Dis.* 18, 502–506.
45. Doughari, H.J., Alois, P.N., Human, I.S. and Benad, S. (2011) The ecology, biology and pathogenesis of *Acinetobacter* spp.: An overview; *Microbes Environ.* 26, 101–112.
46. Fournier, P.E. and Richet, H. (2006) The epidemiology and control of *Acinetobacter baumannii* in health care facilities; *Clin. Infect. Dis.* 42, 692–699.
47. Jawad, A., Heritage, J., Snelling, A.M., Gascoyne-Binzi, D.M. and Hawkey, P.M. (1996) Influence of relative humidity and suspending menstrua on survival of *Acinetobacter* spp. on dry surfaces; *J. Clin. Microbiol.* 34, 2881–2887.
48. Rice, L.B. (2006) Challenges in identifying new antimicrobial agents effective for treating infections with *Acinetobacter baumannii* and *Pseudomonas aeruginosa*; *Clin. Infect. Dis.* 43, 100–105.
49. Loidl, P. (1994) Histone acetylation: facts and questions; *Chromosoma* 103, 441–449.

50. Hong, L., Schroth, G.P., Matthews, H.R., Yau, P. and Bradbury, E.M. (1993) Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA; *J. Biol. Chem.* 268, 305–314.
51. Norton, V.G., Imai, B.S., Yau, P. and Bradbury, E.M. (1989) Histone acetylation reduces nucleosome core particle linking number change; *Cell* 57, 449–457.
52. Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications; *Nature* 403, 41–45.
53. Kuo, M.H. and Allis, C.D. (1998) Roles of histone acetyltransferases and deacetylases in gene regulation; *Bioessays* 20, 615–626.
54. Vetting, M.W., Magnet, S., Nieves, E., Roderick, S.L. and Blanchard, J.S. (2004) A bacterial acetyltransferase capable of regioselective *N*-acetylation of antibiotics and histones; *Chem. Biol.* 11, 565–573.
55. Sampath, V., Liu, B., Tafrov, S., Srinivasan, M., Rieger, R., Chen, E.I. and Sternglanz, R. (2013) Biochemical characterization of Hpa2 and Hpa3-two small closely related acetyltransferases from *S. cerevisiae*; *J. Biol. Chem.* 288, 21506–21513.
56. Williams, K. (1997) Modulation and block of ion channels: A new biology of polyamines; *Cell Signal.* 9, 1–13.
57. Pan, Y.H., Liao, C.C., Kuo, C.C., Duan, K.J., Liang, P.H., Yuan, H.S., Hu, S.T. and Chak, K.F. (2006) The critical roles of polyamines in regulating ColE7 production and restricting ColE7 uptake of the colicin-producing *Escherichia coli*; *J. Biol. Chem.* 281, 13083–13091.
58. Chattopadhyay, M.K., Tabor, C.W. and Tabor, H. (2003) Polyamines protect *Escherichia coli* cells from the toxic effect of oxygen; *PNAS* 100, 2261–2265.
59. Rato, C., Amirova, S.R., Bates, D.G., Stansfield, I. and Wallace, H.M. (2011) Translational recoding as a feedback controller: systems approaches reveal polyamine-specific effects on the antizyme ribosomal frameshift; *Nucleic Acids Res.* 39, 4587–4597.
60. Kwon, D.H. and Lu, C-D. (2006) Polyamines induce resistance to cationic peptide, aminoglycoside, and quinolone antibiotics in *Pseudomonas aeruginosa* PAO1; *Antimicrob. Agents Chemother.* 50, 1615–1622.
61. Shah, P. and Swiatlo, E. (2008) A multifaceted role for polyamines in bacterial pathogens; *Mol. Microbiol.* 68, 4–16.
62. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool; *J. Mol. Biol.* 215, 403–410.
63. Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment; *J. Mol. Biol.* 302, 205–217.
64. Gouet, P., Courcelle, E., Stuart, D.I. and Metz, F., (1999) ESPript: multiple sequence alignments in PostScript; *Bioinformatics* 15, 305–308.
65. Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modelling; *Electrophoresis* 18, 2714–2723.

66. Guerois, R., Nielsen, J.E. and Serrano, L. (2002) Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations; *J. Mol. Biol.* 320, 369–387.
67. Luthy, R., Bowie, J.U. and Eisenberg, D. (1992) Assessment of protein models with threedimensional profiles; *Nature* 356, 83–88.
68. Laskowski, R.A., MacArthur, M.W. and Thornton, J.M. (1998) Validation of protein models derived from experiment; *Curr. Opin. Struct. Biol.* 8, 631–639.
69. PROSA ZWiederstein, M. and Sippl, M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins; *Nucleic Acids Res.* 35, 407–410.
70. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.
71. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR System: A new software suite for macromolecular structure determination; *Acta Crystallogr. D, Biol. Crystallogr.* 54, 905–921.
72. Wolf, L.K. (2009) New software and websites for the chemical enterprise; *Chem. Eng. News* 87, 31–33.
73. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function; *J. Comput. Chem.* 19, 1639–1662.
74. He, H., Ding, Y., Bartlam, M., Sun, F., Le, Y., Qin, X., Tang, H., Zhang, R., Joachimiak, A., Liu, J., Zhao, N. and Rao, Z. (2003) Crystal structure of tabtoxin resistance protein complexed with acetyl coenzyme A reveals the mechanism for beta-lactam acetylation; *J. Mol. Biol.* 325, 1019–1030.
75. Angus-Hill, M.L., Dutnall, R.N., Tafrov, S.T., Sternglanz, S. and Ramakrishnan, V. (1999) Crystal structure of the histone acetyltransferase Hpa2: A tetrameric member of the Gcn5-related *N*-acetyltransferase superfamily; *J. Mol. Biol.* 294, 1311–1325.
76. Dutnall, R.N., Tafrov, S.T., Sternglanz, R. and Ramakrishnan, V. (1998) Structure of the histone acetyltransferase Hat1: A paradigm for the GCN5-related *N*-acetyl transferase superfamily; *Cell* 94, 427–438.
77. Wolf, E., Vassilev, A., Makino, Y., Sali, A., Nakatani, Y. and Burley, S.K. (1998) Crystal structure of a GCN5-related *N*-acetyltransferase: *Serratiamarcescens* aminoglycoside 3-*N*-acetyltransferase; *Cell* 94, 439–449.
78. Peleg, A.Y., Seifert, H. and Paterson, D.L. (2008) *Acinetobacter baumannii*: emergence of a successful pathogen; *Clin. Microbiol. Rev.* 21, 538–582.
79. Maragakis, L.L. and Perl, T.M. (2008) *Acinetobacter baumannii*: epidemiology, antimicrobial resistance, and treatment options; *Clin. Infect. Dis.* 46, 1254–1263.
80. Sullivan, D.R., Shields, J. and Netzer, G. (2010) Fatal case of multi-drug resistant *Acinetobacter baumannii* necrotizing fasciitis; *Am. J. Surg.* 76, 651–653.

81. Iacono, M., Villa, L., Fortini, D., Bordoni, R., Imperi, F., Bonnal, R.J., Sicheritz-Ponten, T., De Bellis, G., Visca, P., Cassone, A. and Carattoli, A. (2008) Whole-genome pyrosequencing of an epidemic multidrug-resistant *Acinetobacter baumannii* strain belonging to the European clone II; *Antimicrob. Agents Chemother.* 52, 2616–2625.
82. Smith, M.G., Gianoulis, T. A., Pukatzki, S., Mekalanos, J. J., Ornston, L.N., Gerstein, M. and Snyder, M. (2007) New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis; *Genes Dev.* 21, 601–614.
83. Lee, C.Y., Delaney, J.C., Kartalou, M., Lingaraju, G.M., Maor-Shoshani, A., Essigmann, J.M. and Samson, L.D. (2009) Recognition and processing of a new repertoire of DNA substrates by human 3-methyladenine DNA glycosylase (AAG); *Biochemistry* 48, 1850–1861.
84. Sakumi, K. and Sekiguchi, M. (1990) Structures and functions of DNA glycosylases; *Mut. Res.* 236, 161–172.
85. Seeberg, E., Clarke, N.D., Evensen, G., Kaasen, I. and Steinum, A.-L. (1986) In: Myrnes, B. and Krokan, H. (eds), *Repair of DNA lesions introduced by N-nitroso compounds*; Norwegian University Press, Oslo, pp. 51–61.
86. Drohat, A.C., Kwon, K., Krosky, D.J. and Stivers, J.T. (2002) 3-methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member; *Nat. Struct. Biol.* 9, 659–664.
87. Friedberg, E.C., Walker, G.C., Siede, W. and Schultz, R.A. (2006) *DNA Repair and mutagenesis*; 2nd ed., ASM Press, Washington, DC.
88. Riazuddin, S. and Lindahl, T. (1978) Properties of 3-methyladenine-DNA glycosylase from *Escherichia coli*; *Biochemistry* 11, 2110–2118.
89. Bhattacharya, S. and Kumar, P. (2012) An *in silico* approach to structural elucidation of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from *Arabidopsis thaliana*: Hints for herbicide design; *Phytochemistry* 73, 7–14.
90. Guerois, R., Nielsen, J.E. and Serrano, L. (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations; *J. Mol. Biol.* 320, 369–387.
91. Singh, D., Sharma, K.K., Dhar, M.S. and Viridi, J.S. (2014) Molecular modeling and docking of novel laccase from multiple serotype of *Yersinia enterocolitica* suggests differential and multiple substrate binding; *Biochem. Biophys. Res.* 449, 157–162.
92. Kalyanamoorthy, S. and Chen, Y.P.P. (2014) Modelling and enhanced molecular dynamics to steer structure-based drug discovery; *Prog. Biophys. Mol. Biol.* 114, 123–136.
93. Hingorani, K.S. and Gierasch, L.M. (2014) Comparing protein folding *in vitro* and *in vivo*: foldability meets the fitness challenge; *Curr. Opin. Struct. Biol.* 24, 81–90.
94. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The protein data bank; *Nucl. Acids Res.* 28, 235–242.

95. Marti-Renom, M.A., Madhusudhan, M.S., Fiser, A., Rost, B. and Sali, A. (2002) Reliability of assessment of protein structure prediction methods; *Structure* 10, 435–440.
96. MacKerell, A.D., Bashford, D., Bellott, Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher III, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wirkiewicz-Kuczera, J., Yin, D. and Karplus, M. (1998) All-atom empirical potential for molecular modelling and dynamics studies of proteins; *J. Phys. Chem. B* 102, 3586–3616.
97. Hooft, R.W., Vriend, G., Sander, C. and Abola, E.E. (1996) Errors in protein structures; *Nature* 381, 272–272.
98. Wallner, B. and Elofsson, A. (2005) All are not equal: A benchmark of different homology modeling programs; *Protein Sci.* 5, 1315–1327.
99. Zhu, X., Yan, X., Carter, L.G., Liu, H., Graham, S., Coote, P.J. and Naismith, J. (2012) A model for 3-methyladenine recognition by 3-methyladenine DNA glycosylase I (TAG) from *Staphylococcus aureus*; *Acta Cryst. F*, 68, 610–615.
100. Krissinel, E. and Henrick, E. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions; *Acta Cryst. D*, 60, 2256–2268.
101. Balasubramanian, P.K., Balupuri, A., Kothandan, G. and Cho, S.J. (2014) *In silico* study of 1-(4-phenylpiperazin-1-yl)-2-(1H-pyrazol-1-yl) ethanones derivatives as CCR1 antagonist: Homology modeling, docking and 3D-QSAR approach; *Bioorg. Med. Chem. Lett.* 24, 928–933.
102. Wolf, L.K. (2009) New software and websites for the chemical enterprise; *Chem. Eng. News* 87, 31–33.
103. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) CHARMM: A program for macromolecular energy minimization and dynamics calculations; *J. Comp. Chem.* 4, 187–217.
104. Banerjee, A., Santos, W.L. and Verdine, G.L. (2006) Structure of a DNA glycosylase searching for lesions; *Science* 311, 1153–1157.
105. Metz, A.H., Hollis, T. and Eichman, B.F. (2007) DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG); *EMBO J.* 26, 2411–2420.
106. Kwon, K., Cao, C. and Stivers, J.T. (2003) A novel zinc snap motif conveys structural stability to 3-methyladenine DNA glycosylase I; *J. Biol. Chem.* 278, 19442–19446.
107. Pawlak, S.D., Radlinska, M., Chmiel, A.A., Bujnicki, J.M. and Skowronek, K.J. (2005) Inference of relationships in the ‘twilight zone’ of homology using a combination of bioinformatics and site-directed mutagenesis: a case study of restriction endonucleases Bsp6I and PvuII; *Nucl. Acids Res.* 33, 661–671.
108. Kaur, N., Khokhar, M., Jain, V., Bharatam, P.V., Sandhir, R., Tewari, R. (2013) Identification of druggable targets for *Acinetobacter baumannii* via subtractive genomics and plausible Inhibitors for MurA and MurB; *Appl. Biochem. Biotech.* 171, 417–436.

109. Choudhury, C., Deva, P.U. and Sastry, G.N. (2014) Molecular dynamics investigation of the active site dynamics of mycobacterial cyclopropane synthase during various stages of the cyclopropanation process; *J. Str. Biol.* 187, 38–48.
110. Cao, C., Kwon, K., Jiang, Y.L., Drohat, A.C. and Stivers, J.T. (2003) Solution structure and baseperturbation studies reveal a novel mode of alkylated base recognition by 3-methyladenine DNA glycosylase I; *J. Biol. Chem.* 278, 48012–48020.



3.1. Introduction

The genus *Acinetobacter* belongs to the family Moraxellaceae and the order Pseudomonadales. Till date, approximately 25 species within the genus *Acinetobacter* have been identified. Still it is difficult to differentiate among different members of this genus. The bacteria in the genus *Acinetobacter* are non-motile, oxidase-nitrate negative, and non-lactose fermenting [1, 2]. The increasing protein sequence data has initiated a move from the traditional protocols of drug discovery to structure-based drug designing. In past decades, infections caused by *Acinetobacter* have been noticed sporadically and which are most common in case of immune compromised patients. Presence of *Acinetobacter* in health care related outbreaks targeting severely ill patients have documented it as an opportunistic, healthcare-associated, multidrug-resistant pathogen. The emergence of the MDR strain of *Acinetobacter* has made the currently ongoing therapies ineffective resulting in the scarcity of novel drugs and drug targets. The Center for Disease Control and Prevention (CDC) *A. baumannii* holds a major share (approx. 80%) of the reported *Acinetobacter* infections [3]. *A. baumannii* is capable of surviving for long periods of time on inert surfaces. This prolonged survival capacity along with the multidrug resistance, colonization potential, and contact transmission are the challenging factors in controlling *Acinetobacter*.

As described in Chapter 1, the enzyme Hpa2 from *A. baumannii* is chosen as an unconventional drug target. Initial reports of Hpa2 from yeast in native and complex (with acetyl-CoA) forms were described by **August-Hill and his co-workers**. They have described different oligomeric states of Hpa2 in solution in presence as well as in absence of its substrate acetyl-CoA. The dimer of this protein (AB) is formed by the extensive interaction between the residues numbered 26 to 156, and it involves direct interactions of ~50 residues. Furthermore, it consists of 42 H-bonding and 411 non-bonding interactions. Substrate acetyl-CoA of Hpa2 has been shown to be affecting its dimeric state developing the tetrameric conformation of the protein. Crystal data of Hpa2-acetyl-coA complex has indicated towards the significance of the interaction of adenine base-pair moieties of the substrate in the formation of tetramer. This change in oligomeric conformation of Hpa2 represents a novel method of protein oligomerization facilitated by the substrate base-pairs (1qsm). Yjcf protein, from *Bacillus subtilis* which belongs to superfamily Gcn5-related *N*-acetyltransferase shares significant sequence similarity with Hpa2. Structural studies on Yjcf show that usually

available as monomeric in the solution (1q2y). Another protein having significant sequence identity with Hpa2 is a putative N-acetyltransferase from *Lactobacillus plantarum*. Crystal structure of this protein (3efa) has shown it to be monomeric in native form, while structure in complex form of this protein is not available. But another protein which is an acyl-CoA N-

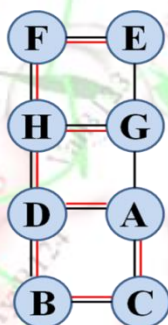


Figure 1. Arrangement of different chains in 1xeb structure.

acetyltrans-ferase from *Pseudomonas aeruginosa*, with significant similarity in sequence with Hpa2 (*A. baumannii*) exists in octameric state in solution in absence of substrate molecules. Crystal data of this protein (1xeb) gives insight about the interaction contributed by different chains in the formation of homo-octameric structure. Details of the interactions in terms of energetics and residues contributing in the interactions are given in Appendix III-1.

Analysis from Appendix III-1 and Figure 1 suggests that the structure of protein pdb id 1xeb consists of eight similar chains, and these eight chains make 10 different types of dimer interactions, four among which are more stable. Analysis of protein interfaces suggests that the BC, AD, EF and GH quaternary structures are stable in solution. The interaction energy and buried surface area for all the four quaternary structures are almost same and hence it can be concluded that the probability of their formation is equal and/or these protein dimers may exist in equilibrium. On the basis of the residues involved in the formation of interface, these 10 possible dimers can be grouped into three categories: First category involves BC, AD, EF and GH quaternary structures involving the residues numbered 10 to 45 and their dimerization is stabilized by ~6 H-bonded and ~110 non-bonded interactions. As their nature as well as interaction is same, these four may be considered as one species. Second category involves AC, BD and DH quaternary structures which involving the residues numbered 73 to 148 and dimer is stabilized by ~3 H-bonded and ~50 non-bonded interactions. Third category involves AG, EG and FH quaternary structures involving residues numbered 56 to 90 and the dimer is stabilized by ~1 H-bonded and ~10 non-bonded interactions and these dimers might be unstable in the solution. According to this analysis, it can be hypothesized that the dimeric conformations of third category might be the transient dimer. The Hpa2 from yeast has been reported to exist in dimeric state while it forms tetramer in the presence of substrate acetyl-

CoA in solution [4]. The three-dimensional structure of the enzyme Hpa2 from different species/sources [1q2y, 1xeb] [4] have shown that the enzyme exists as monomer, dimer, tetramer and as higher oligomeric forms. Higher state of oligomerization might be contributing significantly in functioning and stability of the proteins like Hpa2. Based on the literature information discussed above, it appears that proteins closely related to Hpa2 exist in different oligomeric states and hence here we have performed the experimental studies to analyse the oligomeric state of Hpa2. Primary classification of bacteria is done on the basis of Gram staining. The Gram stain differentiates bacteria into two types, *i.e.*, Gram-positive and Gram-negative. Gram-positive bacteria would appear violet in colour due to the retention of stain by their cell wall. Major diseases caused by bacteria are listed in Table 1 of *Chapter 1*. Gram-negative bacteria would not be able to retain stain and would appear pink under the microscope. This happens because of the difference in composition of their cell wall. The cell wall forms a tight barrier against the outside environment providing immunity to these both bacterial types from different medications as well as host immune responses. Only few human enzymes and some microbial secondary metabolites have been obscured to neutralize this defense barrier of bacteria. First effective metabolite to be discovered against bacterial cell wall is penicillin with the discovery it declared the beginning of golden age of antibiotics. To elucidate the complete 3D structure we also planned to crystallize the protein Hpa2. Before proceeding towards the experimental section, brief description of different biophysical techniques used is given.

3.2. Techniques used for biophysical characterization of Hpa2

3.2.1. Size exclusion chromatography

Size exclusion chromatography, also known as partition chromatography in which the separation of protein molecules takes place based on their molecular size and shape. The stationary phase of the column is made up of the beads having pores that span a relatively narrow size range of molecular dimensions. The bead network retards large molecules and allows small molecules to penetrate through the pores and the molecules that can enter the beads take a convoluted, and therefore, longer path through the column due to which their movement is slow. There is a mathematical logarithmic relation between the volume eluted from the column and the size of the molecule (globule protein). The two assumptions are

taken into consideration during this experiment, (i) molecules are spherical in the shape, and (ii) the protein molecules do not make any interaction with the gel material [5, 6].

3.2.1.1. Determination of the protein molecular mass

Size exclusion chromatography is a reliable method for the determination of relative molecular mass of protein. It can be applied in substitution of native PAGE for determination of protein molecular weight. The void volume is first determined by injecting 50 μ L of 8 mg/mL blue dextran and establishing its elution volume (40 mL). Prior to running the protein of interest, the calibration curve for the same column is prepared by running six standard proteins of known molecular weight (ribonuclease A: 13.7 kDa, carbonic anhydrase: 29 kDa, ovalbumin: 44 kDa, thyroglobulin: 670 kDa, aldolase: 158 kDa and aprotinin: 6.5 kDa). This curve enables the estimation of molecular size value of unknown protein.

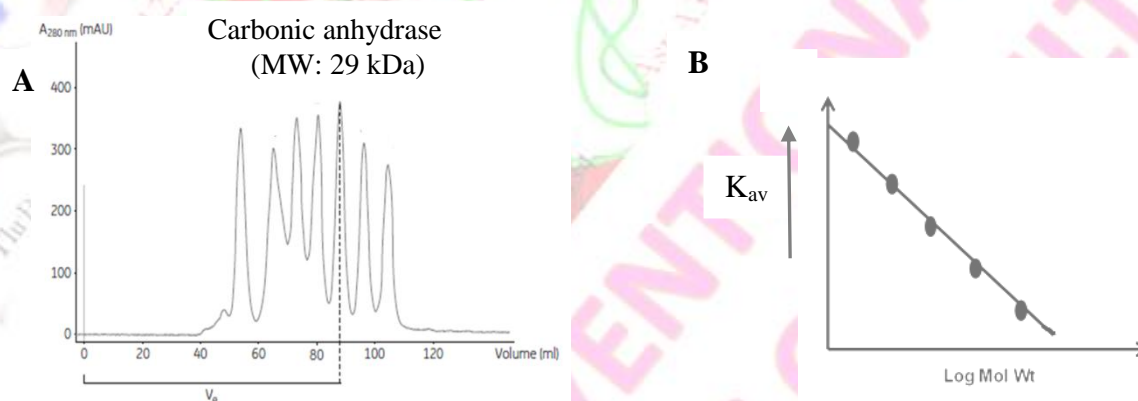


Figure 2. Determination of molecular mass using gel filtration chromatography.

The retention volume for each standard protein was recorded separately and their distribution coefficient (K_{av}) values were calculated using equation,

$$K_{av} = (V_e - V_o) / (V_t - V_o),$$

where, V_e is the elution volume, V_o is the void volume and V_t denotes the total volume. A graph of K_{av} as a function of the log molecular weight produces a standard curve (Figure 2). From this graph the molecular weight of unknown protein was calculated.

3.2.2. Circular dichroism

Circular dichroism is the difference in the absorption of left-handed circularly polarised light (L-CPL) and right-handed circularly polarised light (R-CPL) and it occurs when a molecule

contains one or more chiral chromophores (light-absorbing groups). Circular dichroism of chiral compounds is calculated using the formula,

$$\text{Circular dichroism} = \Delta A(\lambda) = A(\lambda)_{\text{LCPL}} - A(\lambda)_{\text{RCPL}}, \text{ where } \lambda \text{ is the wavelength.}$$

CD spectroscopy is used extensively to study biological macromolecules like proteins, DNA and RNA. A primary use of CD is in analysing the secondary structure or conformation of macromolecules and since the protein structure is sensitive towards its environment (temperature or pH), circular dichroism can be used to observe the impact of environmental conditions on the structure [7–11]. In the present study, the CD spectroscopy is used to measure the secondary structure, folding of recombinant Hpa2 protein and effect of temperature and denaturants on the protein folding-unfolding process as a function of change in the CD signal at a particular wavelength (222 nm).

3.2.3. Isothermal titration calorimetry

The calorimetric technique ITC is frequently used to study the protein–ligand interactions. An ITC experiment involves successive additions of ligand to a solution of protein, present in the reaction cell. After each addition, the formation of a specific amount of protein–ligand complexes will take place, along with release of energy as depicted in Figure 3. The binding affinity can be calculated by monitoring this release or uptake of heat.

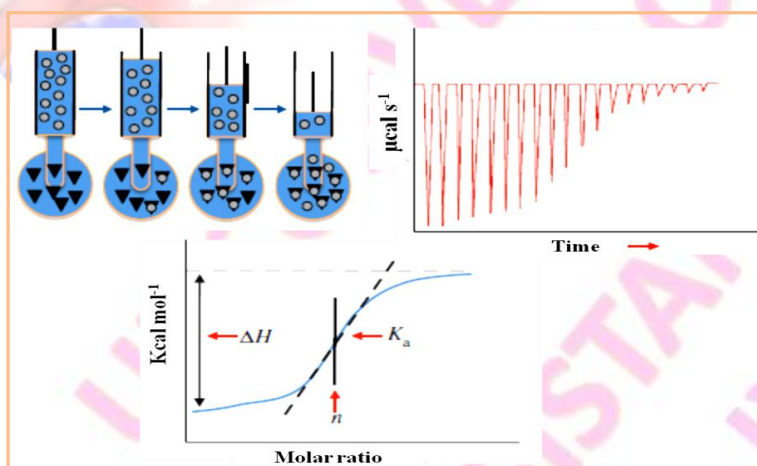


Figure 3. ITC, a characteristic titration experiment with its evaluation.

The ITC provides enthalpy change ΔH , entropy change ΔS , and association binding constant K_a . Then ΔG is calculated from these values using the equation,

$$\Delta G = \Delta H - T\Delta S$$

where $T = 298$. The ITC is considered as an excellent tool to study the thermodynamics of binding. ITC is the only experimental technique where the binding constant (K_d), Gibbs free energy of binding (ΔG), enthalpy (ΔH) and entropy (ΔS) can be determined in a single experiment. The ITC instrument is very sensitive and it can measure the heat changes in the range of 10^{-8} W and association constants, K_a in the range of 10^2 – 10^9 M^{-1} , accurately.

3.2.4. Spectrofluorimetric studies

Fluorescence spectroscopy is considered to be primary research tool in biochemistry and biophysics. In the fluorescence experiment, titrations are carried out and change in fluorescence intensity is recorded.

The enhancement in each case is calculated by

$$\Delta I = I - I_0$$

The aromatic amino acids tryptophan (W) and tyrosine (Y) residues are natural fluorophores present in the proteins. Among both of these, W has the higher quantum yield and its emission spectra are sensitive for its environment and due to which, W is commonly considered as fluorescent probe for the intrinsic fluorescence experiment. It is used to study protein folding/unfolding and dynamics. The λ_{\max} of a buried W is around 335 nm, whereas fully solvent accessible W has λ_{\max} around 355 nm. Therefore, fluorescence is one of the most effective techniques for the conformational changes in the protein on varying the environmental conditions. Since it is a slower process (10^{-9} to 10^{-8} sec) than absorption, during this period of time, all kinds of processes, including protonation or deprotonation reaction, local conformational changes can occur [12–13]. Here the change in the fluorescence of intrinsic protein fluorophores is recorded. The change in wavelength was measured (at 350 nm) on varying the temperature and denaturant concentration.

3.2.5. Native PAGE

The polyacrylamide gel electrophoresis is a high resolution separation technique, which separates protein molecules on the basis of charge, size and conformations. The key advantage of this technique is that the biological activity of protein remains unaffected. Hence this

technique facilitates the estimation of molecular weight of protein in native state. However, in the SDS-PAGE, the protein molecule is denatured and separation takes place on the basis of the size. The standard protocol for native PAGE as described by Arndt and co-workers [14] has been followed. Although molecular weight determination using native PAGE is not reliable, mobility of protein in the native PAGE is influenced by both charge and size.

3.2.6. Glutaraldehyde crosslinking assays

The glutaraldehyde crosslinking is a simple and reliable method for determination of oligomeric state of the protein. It is a very simple method requiring mixing of pure glutaraldehyde solution with the protein solution. Glutaraldehyde links covalently to the amine groups of lysine or hydroxylysine in the protein molecules generating a structure very stable than the structure obtained by the physical aggregation of protein molecules (Figure 4). Here glutaraldehyde crosslinking experiment was performed by following the method described in the protocol [15]. The steps involved in the assay are described here; the protein was purified in Tris buffer and after purification, the Tris buffer was exchanged with 50 mM sodium phosphate buffer of pH 7.6. The buffer exchange was done to remove the amino groups present in the Tris buffer as they react with glutaraldehyde, affecting the outcome of the reaction *i.e.*, between protein and glutaraldehyde (Figure 4). In the present study, the experiment was performed using 24-well crystallization plates and a siliconized cover slip as depicted in Appendix III-2 [15]. To achieve the crosslinking of protein, 40 mL of 12.5% glutaraldehyde solution (v/v) acidified with 1 mL 5 N HCl was placed in the bottom of the well. After that, 15 μ L of protein solution was placed on to the cover slip, and the cover slip was placed on the reservoir in the inverted manner. The reservoir was sealed with vacuum grease and entire setup was incubated for 15 min at 37 °C, then the sample was mixed with an equal volume of 2X SDS-PAGE loading buffer and boiled using dry bath for 4 min. Cross-linked oligomers were analyzed on 15% SDS-PAGE followed by Coomassie Blue R-250 staining and destaining.

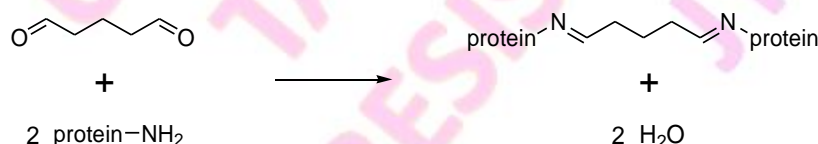


Figure 4. Crosslinking reaction of glutaraldehyde with free $-\text{NH}_2$ group present in protein.

3.3. Materials

3.3.1. Chemicals

All the chemicals used were of analytical grade and most of the chemicals were purchased from Fluka (Buchs, Switzerland), Sigma-Aldrich (Deisenhofen, Germany), Merck (Dramstadt, Germany), and Himedia. Kits for the PCR purification, gel extraction and plasmid miniprep were purchased from Qiagen. Unit for agarose gel electrophoresis, gel imager were from Bio-rad and chemicals used in the DNA gel electrophoresis such as ethidium bromide, agarose, and gel loading dye were purchased from Biochem. Isopropyl β -D-1-thiogalactopyranoside used to induce protein expression was purchased from Biochem. The reagents used in the sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE) and in the Bradford assay were purchased from the Bio-rad. Columns for gel filtration; HisTrap HP and HiLoad 16/600 Superdex 200 were purchased from GE healthcare. The amicon centriprep concentrators with 3 kDa cut-off and Spectra/Por dialysis tubing MW 3,500 were purchased from the Millipore.

3.3.2. Bacterial strains and vectors

Escherichia coli DH5 α and BL21 (DE3) strains and cloning and expression plasmids were purchased from Novagen.

3.3.3. Molecular biology enzymes

Restriction endonuclease enzymes (*Nde*I, *Xho*I), bovine serum albumin, reaction buffers, T4 DNA ligase, DnaseI, Rnase A, Pfu Turbo polymerase and Taq polymerase were purchased from New England Biolabs (NEB). Bovine serum albumin (BSA) and reaction buffers were obtained from NEB. Enzymes utilized during purification, lysozymes and deoxyrobonuclase I (DNase I) were purchased from the Sigma-Aldrich.

3.3.4. Oligonucleotides

DNA oligonucleotides used in the PCR amplification were designed with the help of oligoanalyzer tool of IDT technologies. After checking that *Nde*I and *Xho*I sites are absent in the desired gene sequence, these R.E. sites were introduced, respectively, in forward and reverse primers. The DNA oligonucleotides were purchased from Biolinkk, Bangalore, India.

The sequences of primers presented in 5' to 3' orientation. The synthesis of gene coding for the Hpa2 protein of *A. baumannii* were done from the Biolinkk.

3.3.5. Culture media

Luria–Bertani (LB) broth and LB agar for growing *E. coli* cells were purchased from Merck speciality chemicals as dried granules. The LB agar consists of 1% [w/v] tryptone, 0.5% [w/v] yeast extract, 1% [w/v] NaCl. The LB culture media was sterilized by autoclaving at 121 °C for 15 min at 15 psi pressure. Antibiotics – kanamycin and chloramphenicol were purchased from Himedia chemicals, India, and added into culture for the selective growth. Antibiotics: kanamycin (Stock solution: 100 mg/mL in *ddH*₂O, working concentration 50 µg/mL), chloramphenicol (stock solution: 100 mg/mL in absolute ethanol, working concentration: 35 µg/mL).

3.3.6. Solutions of crystallization

For the crystallization purpose, well known commercial screens *i.e.*, PEG ion I & II, crystal screen I & II, salt, index and crystal screen cryo were purchased from the Hampton Research. For manual optimization of conditions, chemicals were purchased from Sigma Aldrich. The solutions were prepared using high grade nuclease-free water and then filtered through 0.22 µM filter [from Millipore India]. The reagents were prepared and stored at 4 °C.

3.4. Methods

3.4.1. Cloning of *Hpa2* gene

The coding region of *Hpa2* gene encoding the putative Hpa2 protein was PCR amplified using forward and reverse primers containing *Nde*I and *Xho*I restriction sites (underlined) 5' GAT TCT CAT ATG ATG AAA ACT CAA CGC TGC GGA TG 3'(forward) and 5' GAT TCT CTC GAG TTA TGA AGC TTT AAA TTG ACA ATC ATT 3' (reverse). The amplified fragments were subjected for purification using Qiagen PCR purification kit, to remove the unused dNTP's or primer-dimers. The isolated plasmid and purified PCR product were double digested at 37 °C for 1 h with both the restriction enzymes. The digested plasmid and PCR products were then loaded on to low melting agarose gel and after running the digested plasmid, and PCR products were extracted from gel using Qiagen gel extraction kit. After

extraction, the product is ligated into pET28a vector in between *Nde*I and *Xho*I sites with a 6xHis-tag at N-terminal. For tag removal, TEV protease site is present in between 6xHis-tag and *Hpa2* gene. The ligated product was transformed into XL-1 blue cells, which were grown on the LB agar plates containing 50 µg/mL of kanamycin and incubated at 37 °C for overnight. The next day, colonies were screened for the clone using PCR amplification and restriction enzyme digestion. Screened plasmids were sent for sequencing and after confirmation it was used for *Hpa2* protein expression. The expression vector pET28-*Hpa2* possesses T7 promoter, which is essential for the gene transcription during expression in *E. coli*. The expression construct has an N-terminal 6xHis-tag which assists in the purification of the protein of interest with immobilized metal ion affinity chromatography (IMAC) columns. General steps of the experimental work done are depicted in **Figure 5**.

3.4.2. Optimization of *Hpa2* expression

After confirmation of cloning pET28-*Hpa2* plasmids expression vector was transformed into *E. coli* BL21 (DE3) cells (using heat shock method). After transformation, *E. coli* BL21 cells were spread on the LB agar plates containing 50 µg/mL of kanamycin and 35 µg/mL of chloramphenicol, and incubated at 37 °C for overnight. Single colony from the transformed plate was used to inoculate 10 mL of LB containing 50 µg/mL kanamycin, and 35 µg/mL chloramphenicol, in a 100 mL culture tube. 100 µL of this overnight grown culture was used to inoculate fresh 10 mL of LB. The inoculated culture was incubated at 37 °C at 200 rpm until the optical density at 600 nm (O.D. 600) reached to 0.6–0.7. The culture was then induced with different concentrations of IPTG (0.25–1.0 mM). After induction, the culture was grown at 18, 25 and 37 °C temperatures for 16, 7 and 4 h, respectively. After incubation, cells were pellet down at a speed of 6000 rpm for 10 min, the pellets were then resuspended in 5 mL of 50 mM Tris buffer [pH 7.6] and lysed using an ultra sonicator [10 sec pulse on, 10 sec pulse off, 50% of amplitude] the sample was kept in ice bucket to avoid overheating. The disrupted cells were transferred into the centrifuge tubes, and spun down at 10000 rpm for 45 min, at 4 °C. The soluble part [supernatant] and insoluble fractions [pellet] were separated and collected in separate falcons. Expression and solubility of this sample was analysed in presence of control (uninduced sample) on SDS–PAGE after mixing with 6xSDS loading buffer.

SDS–PAGE under reducing and denaturing conditions can be used to separate proteins based on mass. Samples were prepared for electrophoresis by mixing an 2–20 μg of protein, with 6x SDS–PAGE load buffer [0.125 M Tris buffer pH 6.8, 9% β -mercaptoethanol (β ME), 20% glycerol, 4% SDS, 0.1% bromophenol blue]. Samples were heated at 98 $^{\circ}\text{C}$ for 5 min and then loaded on 15% SDS–PAGE gels. Broad Range (10–270 kDa) protein ladder (5 μL) from GeneDireX was used as standard for each gel. Gels were run according to the Lamelli method and after completion, gel was stained for 1–2 h using CBB R-250 dye. Gel was then destained for 45 min, using the destaining solution.

3.4.3. Purification of Hpa2

Homogeneously pure Hpa2 was achieved after employing a three-step strategy of purification. Initial purification step employed affinity-based employed affinity based chromatography using Ni^{2+} beads. His-tag removal step of purification was performed by cleavage mediated by TEV (Tobacco Etch Virus) protease at its recognition site. Finally, size exclusion chromatography was done to produce the pure stable protein suitable for crystallization and other biophysical studies. Detailed description of the purification strategy is given below:

The 1 liter of LB broth media containing suitable concentrations of antibiotics was inoculated with 10 mL of overnight grown culture [having transformed cells]. This culture was incubated at 37 $^{\circ}\text{C}$ with 200 rpm shaking until the optical density at 600 nm (O.D. 600) reached to 0.6–0.7. At this stage, protein expression was induced by addition of 0.5 mM IPTG, followed by incubation at 18 $^{\circ}\text{C}$ at 200 rpm for 16 h. The cells were harvested by centrifugation at 5000 rpm for 10 min at 4 $^{\circ}\text{C}$ and stored at -80°C . For protein purification, cell pellet was resuspended in 20 mL of binding buffer [50 mM Tris buffer pH 7.6, 250 mM imidazole, 5% glycerol, 250 mM NaCl, 5 mM DTT]. The cells were incubated in presence of 0.1 mM PMSF for 10 min and then lysed using French press [Constant cell Disruptor Systems, UK]. The disrupted cells were centrifuged at 10500 rpm, 60 min, 4 $^{\circ}\text{C}$ to separate the soluble portion from the insoluble fraction. In the meantime the Ni-bead column was equilibrated with the binding buffer. The cleared cell lysate was then loaded on the equilibrated column and incubated with Ni beads for 60 min at 4 $^{\circ}\text{C}$. The protein was eluted from column using elution buffers with the gradient concentrations of imidazole [50 mM Tris buffer pH 7.6, 5% glycerol, 250 mM NaCl, 5 mM DTT and 100, 150, 250 mM imidazole], followed by washing of column by 1 M imidazole. Fractions of 5 mL volume were collected over the

course of gradient wash. All the fractions were analysed by 15% SDS–PAGE gel to confirm the presence of protein as well as the sample. Fractions containing protein were dialysed in dialysis buffer [50 mM Tris buffer pH 7.6, 5% glycerol, 100 mM NaCl, 5 mM DTT], and TEV protease was added to remove the N-terminal 6xHis-tag at 4 °C. The TEV cleaved protein was then reloaded on the equilibrated column to separate the histidine-tag cleaved protein from uncleaved one. The protein fractions were collected and the cleaved Hpa2 protein was then concentrated using Amicon [3 kDa cut-off] till 1 mL volume [conc. ~ 3 mg/mL] and loaded on pre-equilibrated [50 mM Tris buffer pH 7.6, 5% glycerol, 0.5 mM DTT] HiLoad 16/600 Superdex 200 column to remove impurities further. The protein eluted from Superdex column was then concentrated using Amicon Ultra-15 centrifugal filter device [3 kDa cut-off], only after confirmation on 15% SDS–PAGE. Protein was concentrated to around 10 mg/mL and was stored at –80 °C after flash freezing with liquid nitrogen till further use.

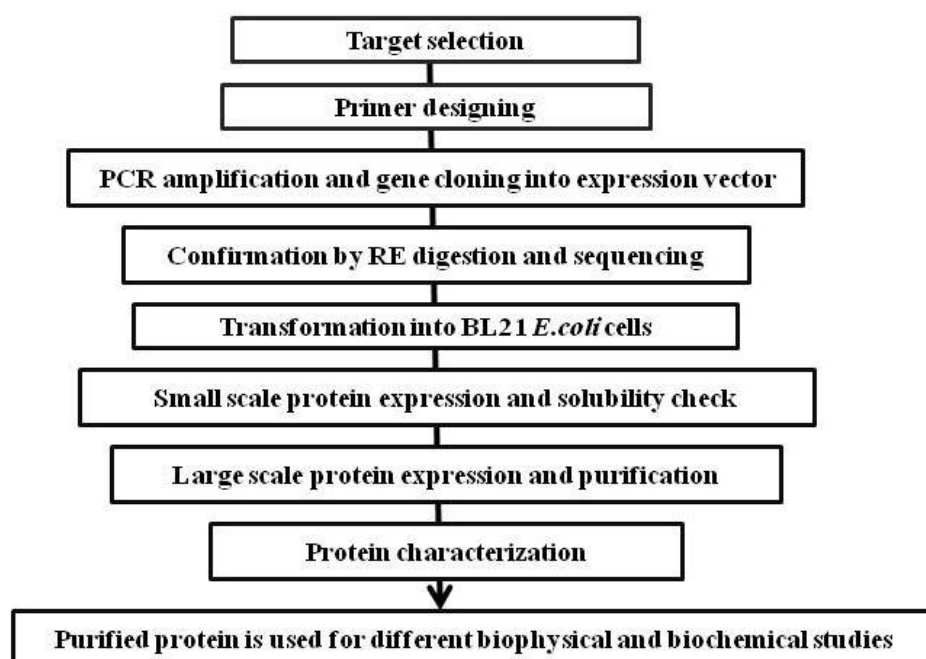


Figure 5. Flow chart showing standard process of Hpa2 protein, study from primer designing to the characterization of protein.

3.4.4. Determination of Hpa2 protein concentration

The Hpa2 protein concentration was determined using Bradford reagent provided by Bio-Rad [16]. Assay was performed as per the instruction provided by the supplier. The 5 μ L of

concentrated protein sample was mixed with 795 μL of purification buffer. The 200 μL of Bradford reagent was added to make final volume 1 mL and incubated for 1 min. The protein concentration was determined using Perkin Elmer Lambda 25 UV-visible spectrophotometer by interpolating the absorbance reading at 595 nm into a BSA standard curve.

3.4.5. Analytical size exclusion chromatography

The analytical size exclusion chromatography is used for the determination molecular weight and oligomeric state of Hpa2 protein. As discussed in the *Techniques Section* (3.2.1.1), prior to running the protein of interest, the calibration curve for the size exclusion column is prepared by running standard proteins of known molecular weight. This standard curve enables the estimation of molecular size of Hpa2 protein. The determination of molecular weight of protein Hpa2 was performed at physiological conditions and in the presence of various substrates (acetyl-CoA and polyamines).

The HiLoad 16/600 Superdex 200 column was prepared with running buffer (50 mM Tris buffer pH 7.6, 100 mM NaCl, 5% glycerol, 0.5 mM DTT) at 4 °C, flow rate of 0.5 mL/min. All five standard proteins (1 mg of each) were dissolved in 1.0 mL of the running buffer. Each protein was loaded individually and retention volume for each standard protein was then recorded separately. The distribution coefficient (K_{av}) value for each standard protein was calculated as given in the *Techniques section* (3.2.1.1). A plot of K_{av} as a function of the log molecular weight gives the standard curve. The 500 μL of purified Hpa2 protein (5 mg/mL in the running buffer) was uploaded in the column and run under the conditions as described above. Retention volumes of the peaks were recorded and the MW of each protein was calculated. The same procedure was used for the determination of MW of Hpa2 in native state as well as in the presence of various substrates, denaturant and salt. Prior to injection, all samples were filtered through 0.22 μm filters.

3.4.5.1. Effect of urea and NaCl on Hpa2 oligomerization

The oligomeric state of the Hpa2 is determined in the presence of various concentrations of salt and denaturing agent. For this, freshly purified Hpa2 protein was pre-incubated with variable concentrations of denaturant and salt (2 M and 4 M of urea and NaCl separately) at 4 °C for 4 h. Then protein was loaded onto HiLoad 16/600 Superdex 200 gel filtration column.

Prior to loading of proteins, the gel filtration column was equilibrated with the buffer containing urea or NaCl used during the incubation and elution profiles were analyzed.

3.4.5.2. Effect of substrates binding on the Hpa2 oligomeric state

The oligomeric state of Hpa2 was also analysed in the presence of various substrates and inhibitors. The protein was incubated with the substrates – acetyl-CoA [2.5 mM] and polyamines [3 mM] for 4 to 6 h at the 4 °C, then loaded on HiLoad 16/600 Superdex 200 gel filtration column equilibrated with same buffer and elution profile was analyzed.

3.4.6. Circular dichroism studies

Circular dichroism spectroscopy was performed on Chirascan CD spectrometer (Applied Photophysics, UK). Spectropolarimeter was equipped with a programmable peltier junction temperature controlled cell holder at 25 ± 0.5 °C. All spectra were acquired with continuous purging of dry nitrogen gas in the sample chamber to prevent condensing of moisture on the optical windows. Far-UV CD spectra were obtained at 20 °C from protein samples (1 mg/mL) in 25 mM sodium phosphate buffer pH 7.6 having NaCl (50 mM) and glycerol (5%, v/v). The spectra were recorded in a 1 mm quartz cell at bandwidth 1 nm and time per point 0.5 sec with four repeats. Three scans for each protein sample were accumulated (190–260) to get the average spectrum then base line spectrum was subtracted and thus obtained final spectrum was analyzed by CDSSTR Method [17–18]. In this method, initially, a large database of standard spectra from the proteins of known secondary structures is created. After that secondary structure is calculated using the singular value decomposition. Then superior fits of globular proteins was done. CD titration was also performed at different temperatures (20–75 °C), and values for the CD data are represented in delta epsilon.

3.4.7. Isothermal titration calorimetry binding assay

To determine the binding parameters of Hpa2 with the substrates (acetyl-CoA and polyamines), isothermal titration calorimetry assay were performed using an ITC200 Microcal calorimeter (GE Healthcare). The 225 μ L of 300 μ M pure Hpa2 protein (25 mM Tris buffer pH 7.6), loaded in the calorimetric cell, was titrated against 3 mM polyamine, 2.5 mM acetyl-CoA and 2.5 mM of antibiotics solution using a 299 μ L syringe. About 20 injections of 2 μ L ligand each were injected into the cell at a stirring speed of 2000 rpm. The time interval

between two injections was set 120 sec to ensure optimum mixing and measurement of change in reaction enthalpy. Control experiments were performed by titrating substrate with buffer and this titration data of ligand with protein-free buffer was used as a reference. The resulting thermogram were analyzed using single set of binding sites model to provide the binding affinity (K_a), binding stoichiometry (N), and enthalpy of binding (ΔH). The binding free energy (ΔG) and the entropic contribution to the binding ($T\Delta S$) were then calculated from standard thermodynamic relationships as described [19]. The data analysis was done using Origin software version 7.0.

3.4.8. Spectrofluorimetric studies

Fluorescence spectra were recorded on Fluorolog-3 Spectrofluorimeter LS55 (make HORIBA Jobin Yvon Spex®). Fluorescence emission spectrum was recorded in the range of 295–450 nm upon excitation at 280 nm in steps of 2 °C over a 25–50 °C temperature range. The spectra were also recorded for 5 °C steps over the 50–80 °C range. The slit widths were typically 5 nm and volume of the sample for analysis were varied depending upon the concentration of protein required for achieving satisfactory signal to noise ratio. Sample cells were preheated at each temperature for 90 sec before collecting the data. After smoothing the raw spectra, we plotted the intensities of the emission maxima against the temperature. Titrations were carried out and change in fluorescence intensity was recorded. The enhancement in each case was calculated by

$$\Delta I = I - I_0$$

During protein folding-unfolding studies, the shift in the λ_{\max} was noticed so the graph is plotted as of F_{350}/F_{330} vs. temperature, to avoid the error in the interpretation of results.

3.4.9. Native PAGE

Native PAGE was performed for various concentrations of Hpa2. Different concentrations of Hpa2 were subjected to non-reducing environment for determining the impact of concentration on the oligomeric state of the protein. Electrophoresis was performed by using Tris-glycine buffer pH 8.3 as electrode buffer at a constant voltage of 20 mA at 4 °C for 4 h. Subsequently, the gel was stained by Coomassie Brilliant Blue R-250.

3.4.10. Glutaraldehyde crosslinking assays

Hpa2 protein was purified in Tris buffer, and after purification, the protein was exchanged with 25 mM sodium phosphate buffer pH 7.6 for crosslinking studies. For crosslinking, 40 mL of 12.5% glutaraldehyde solution (v/v) acidified with 1 mL 5 N HCl was added in the well. Then 15 μ L of protein solution (1 mg/mL) was placed onto the cover slip, which was placed on the cover of the well in inverted manner. The entire setup was incubated for 15 min at 37 °C. The cross-linked oligomers were analyzed on 15% SDS–PAGE followed by Coomassie Blue R-250 staining and destaining.

3.5. Results and Discussion

3.5.1. Cloning and overexpression of pET28a-Hpa2

In the *Acinetobacter baumannii* ACICU genome, Hpa2 protein is encoded by the ACICU_RS15490 gene. To overexpress the gene encoding, the protein Hpa2 was cloned into an *E. coli* based expression vector, pET28a with N-terminal 6xHis-tag (Figure 6). The gene is cloned between the *Nde*I and *Xho*I restriction sites. Analysis of PCR product on 1% agarose gel electrophoresis confirmed the presence of a single amplification band of approximately 500 bp size, corresponding to the *Hpa2* gene size (417 bp) (Figure 7A). Plasmid and PCR products were double digested with *Nde*I and *Xho*I restriction enzymes and T4 DNA ligase was used to ligate the digested PCR product into the digested vector. The ligated product was successfully transformed into *E. coli* DH5 α cells. To confirm whether the gene is cloned into vector, the plasmid DNA was isolated from transformed cells and subjected to double digestion with *Nde*I and *Xho*I which yielded the fragment of desired *Hpa2* gene length (417 bp) confirming the cloning (Figure 7B).

The automated DNA sequencing from the company further confirmed the cloning of the *Hpa2* gene in pET vector. In order to optimize the Hpa2 expression, the recombinant vector was transformed into *E. coli* BL21 (DE3) cells successfully using the method described in *Methodology Section* (3.4.1). The small scale expression of Hpa2 was analyzed at seven different IPTG concentrations (0.1 mM to 0.7 mM) as well as different conditions of temperatures (18, 25 and 37 °C). The optimum expression and solubility was achieved with 0.6 mM IPTG concentration at 18 °C temperature (Figure 8).

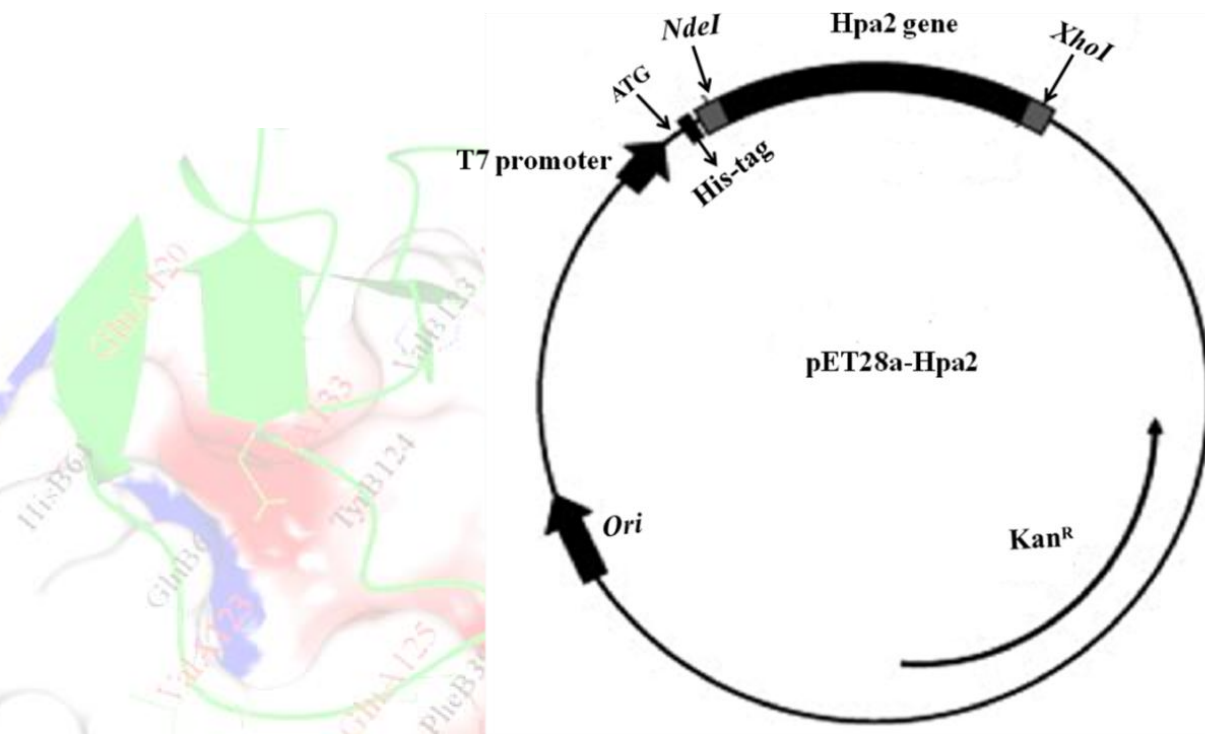


Figure 6. Map of Hpa2 expression plasmid. The open reading frame of *Hpa2* gene was inserted into expression vector pET-28a(+) within *NdeI* and *XhoI* sites.

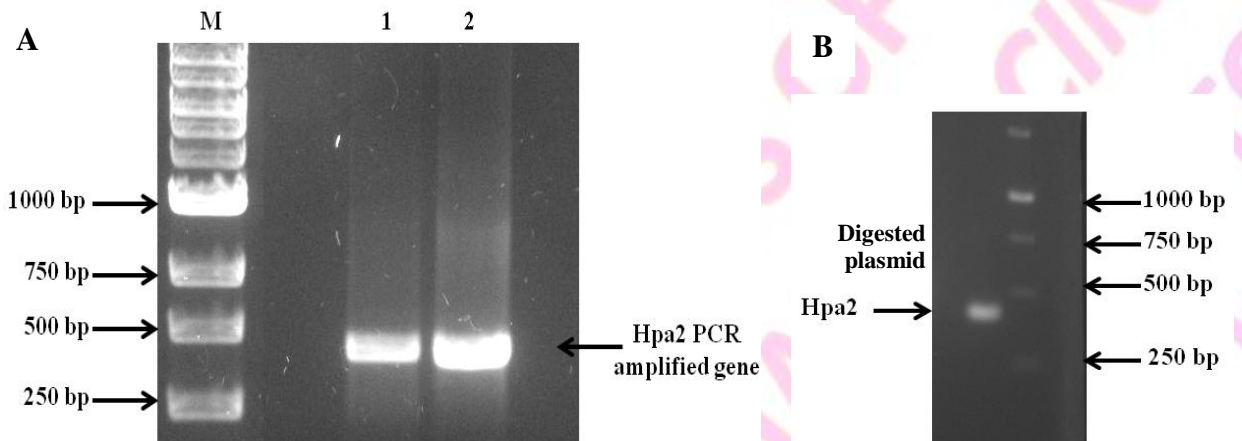


Figure 7. Amplification and cloning of *Hpa2* gene, agarose gel electrophoresis. (A) PCR products. Lane M: DNA marker; Lanes 1 and 2: Amplified PCR product (B) Recombinant pET28a plasmids digested with *NdeI* and *XhoI*. Lane M: DNA marker; Lane L1: pET-28a (+)-*Hpa2* from the screened colony.

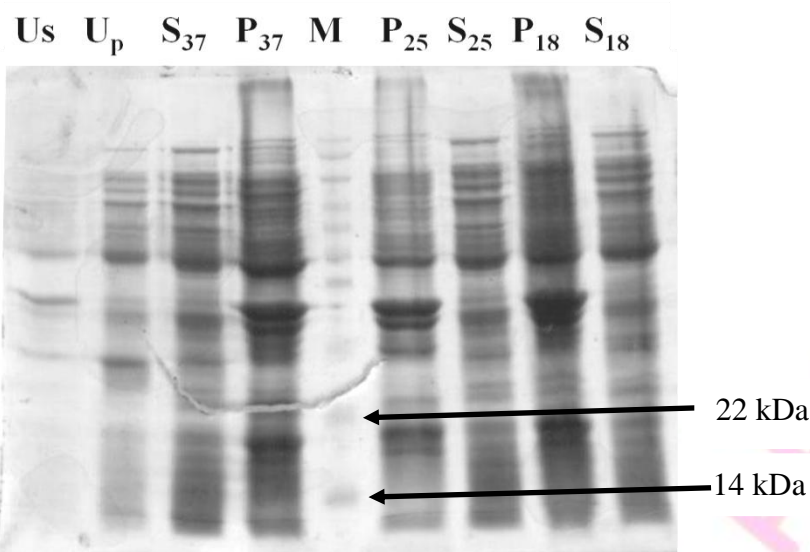


Figure 8. Optimization of Hpa2 Expression, (15% SDS-PAGE analysis). Lane Up: Un-induced pellet, P₁₈ : Pellet at 18 °C, P₂₅ : Pellet at 25 °C, P₃₇ : Pellet at 37 °C, Us: uninduced supernatant, S₁₈ : Supernatant at 18 °C, S₂₅ : Supernatant at 25 °C, S₃₇ : Supernatant at 37 °C.

3.5.2. Purification of Hpa2 protein

Induction of *E. coli* BL21 (DE3) cells harbouring pET-28a(+)-Hpa2 vector with IPTG resulted into high level expression of Hpa2 protein. Purification of Hpa2 protein was done from 1 litre cell pellets. After gentle treatment with lysozyme, the cells were lysed using French press. With the combination of metal affinity-based chromatography (Ni-NTA agarose) and size exclusion chromatography Hpa2 was purified up to more than 95% purity and homogeneity. The SDS-PAGE profile of sample from each stage of the purification is shown in [Figure 9](#). Single preparation from 1 liter of induced culture typically yields 10–12 mg of Hpa2 protein. Highly pure Hpa2 protein was obtained after the gel filtration and all the fractions having protein were run on the 15% SDS-PAGE gel to check the purity and homogeneity ([Figure 12C](#)) of the recombinant Hpa2 protein (~16 kDa size). Protein fractions were incubated with the TEV protease (to remove the 6xHis-tag) and then protein was dialyzed. The tag removed protein was then concentrated to 1000 μ L (5 mg/mL conc.) and purified further with size exclusion chromatography. This purified protein was further analyzed on 15% SDS-PAGE and pure fractions were concentrated to 500 μ L (10 mg/mL conc.) to be used for crystallization. Bradford method [20] was utilized to estimate the Hpa2 protein concentration

using Bradford reagent provided by Bio-Rad Laboratories, USA. The Hpa2 concentration calculated by interpolating the absorbance reading into a standard BSA curve was found to be approximately 10 mg/mL.

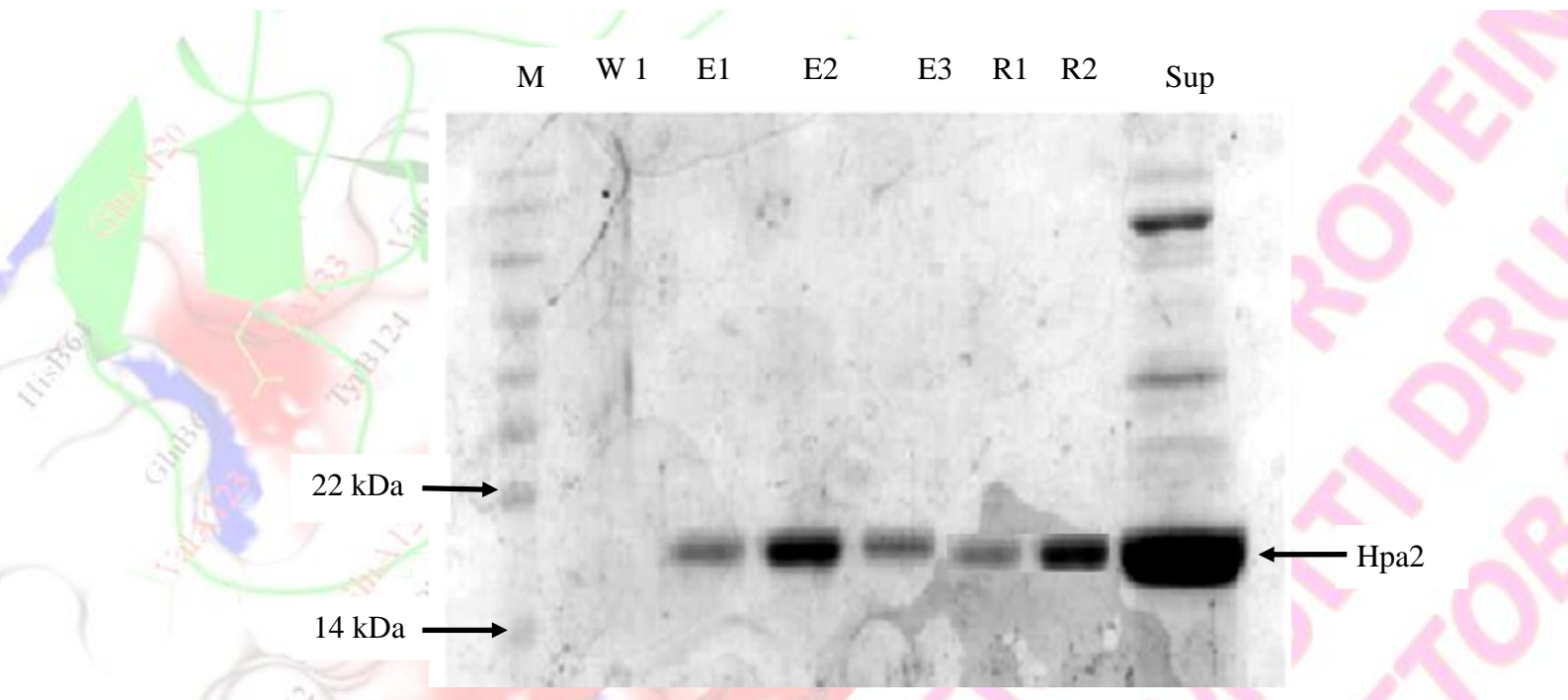


Figure 9. Affinity purification of recombinant Hpa2 [with 6xHis-tag]. Lane M: prestained protein ladder; Lanes E1–E3: purified protein fractions. Reverse NTA step of purification of Hpa2 after the cleavage of 6xHis-tag using TEV protease. Lanes R1 and R2: Hpa2 fractions after the 6xHis-tag cleavage and Sup is supernatant.

3.5.3. Crystallization of Hpa2

Commercial sparse matrix screens from Hampton research, USA were utilized for crystallisation. Each screen had 96 different solutions varying in precipitant, salt and additives combination. Different buffers were used to screen a wide range of pH [pH 4.0–10.0]. The concentration of Hpa2 was kept at 10 mg/mL for the initial screening. Different screens were used for initial screening of crystallization conditions for Hpa2.

The crystallization trial of Hpa2 native and in presence of substrate was carried out separately. In the case of apo-Hpa2, some small needle like crystals with three-dimensional shapes were observed (Figure 10) in condition consisting 0.1 M ammonium acetate, 0.1 M



Figure 10. Crystals of Hpa2 obtained by the hanging-drop vapour-diffusion method using consisting 0.1 M ammonium acetate, 0.1 M Bis-Tris buffer pH 5.5 and 1% polyethylene glycol 8,000, at 4 °C.

Bis-Tris buffer pH 5.5 and 1% polyethylene glycol 8,000. The conditions were further optimized for obtaining diffraction quality crystals of Hpa2.

3.5.4. Glutaraldehyde crosslinking assay

In crosslinking assay, a bifunctional compound glutaraldehyde is used for the chemical modifications of proteins and polymers. Glutaraldehyde crosslinking has wide application in various fields of science for example histochemistry [21–23], microscopy [24, 25], cytochemistry [26], leather tanning industry [27, 28], enzyme technology [29–33], chemical sterilization, biomedical and pharmaceutical sciences [34–35]. In the present study, crosslinking assay was performed to obtain the preliminary information about the oligomeric state of Hpa2 protein.

Solid aggregate formation occurs in presence of glutaraldehyde by non-covalent bonding resulting in strengthening of tertiary structures stabilising the protein against any denaturing environment. In the absence of crosslinking agent, protein Hpa2 migrates as a single band at ~16 kDa (Figure 9) suggesting that either SDS completely dissociates probable oligomeric states of protein or it may exist as monomer in solution. While in glutaraldehyde crosslinking, the monomeric form of Hpa2 is considerably reduced. The major form at positions between 29–43 kDa (Figure 11, Lane 0.05–0.25 mM of glutaraldehyde) represents the dimeric state of Hpa2. In addition, small amount of a tetrameric form at 60 kDa was detected, the intensity of which varies with concentration (Figure 11, Lane 0.30–0.50 mM of

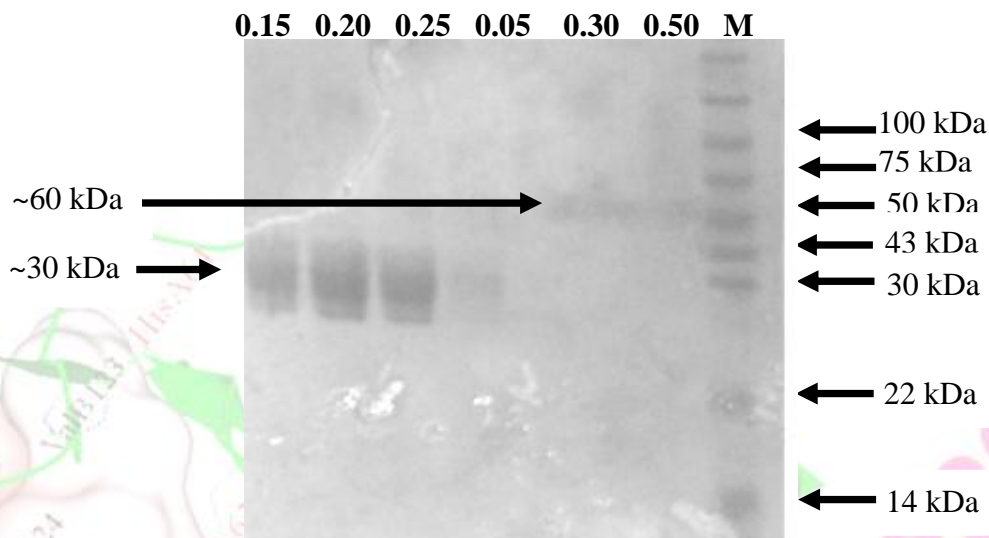


Figure 11. Glutaraldehyde cross-linking assay of Hpa2 protein, for the identification of quaternary protein structure. 15% SDS–PAGE, the Hpa2 protein treated with glutaraldehyde (different concentrations) as described in the protocol.

glutaraldehyde). It is well known from the literature that oligomerization events are dependent on the crosslinking conditions and Hpa2 was found to oligomerise up to tetrameric complexes although it exists as dimer preferentially [4].

3.5.5. Analytical size exclusion chromatography

Size exclusion chromatography used to separate the molecules on the basis of both mass and shape. The molecular weight of the globular protein can be inferred from the elution volume after calibrating the gel filtration column with the standard globular proteins. The oligomeric state of Hpa2 was determined after loading on pre-equilibrated 16/600 Superdex 200 column. Column was calibrated using six globular proteins, spanning a molecular weight (MW) ranging from 6.55 to 670 kDa (Figure 12B). Hpa2 was eluted as a single peak at a volume of 89.6 mL (Figure 12A). The fractions from the column were loaded on 15% SDS–PAGE gel were found to be very pure (Figure 12C). The apparent MW of each standard protein was estimated by interpolation using the globular protein calibration curve (Figure 12D). Based on this standard curve of the molecular weight markers, the molecular weight of the Hpa2 elution peak was calculated. Mass of Hpa2 calculated based on this information was double to the mass calculated on the basis of sequence.

This attests that Hpa2 is migrated as if its molecular weight was approximately twice to that of their actual molecular weight. The fractions obtained from size exclusion chromatography were also analyzed on the native polyacrylamide gels, obtained results verifies the existence of the dimeric Hpa2. (*Native PAGE Section 3.5.6*).

3.5.5.1. Effect of urea and NaCl on Hpa2 oligomerization

The oligomeric state of Hpa2 was also analysed in presence of salt and denaturing agent. The behavior of Hpa2 in presence of such agents has not been studied till now and no information about the impact of concentration of such agents on the stability of Hpa2 is available. In addition, lack of information about the behavior of dimeric Hpa2 in presence of such agents prompted us to examine the effect of different concentrations of NaCl and urea on Hpa2. Incubation of freshly purified protein sample with high NaCl concentration (2 M and 4 M) resulted in the partial dissociation of dimeric Hpa2 enzyme to monomeric state (**Figures 13A and 13B**).

— Generally, protein solutions consisting 50–150 mM salt are used to control pH, ionic strength and osmolality of solution. Since proteins are considered as most complex colloidal system in terms of surface charge, surface chemistry, and size, it needs 50–150 mM of salt to maintain its integrity. Mostly protein surfaces are heterogeneously composed of positive and negative charged, polar and non-polar amino acid residues. Intermolecular or intramolecular interactions of protein molecules can have different origins such as electrostatic, hydrophobic, van der Waals, and hydrogen bonding [36–39]. It is not easy to identify the accurate relative contributions of each type of interactions towards the overall protein–protein interactions. The presence of NaCl in buffer is known to influence the elution volume of protein by altering their hydrodynamic radius. Due to this reason, it is necessary to modify the buffer composition having same salt concentration as in the protein sample to avoid the nonspecific interactions in the column and to reclaim separation based on molecular size alone due to which we used the same amount of salt in the buffer too.

Analysis of elution profile at NaCl concentrations 2 M and 4 M has shown two peaks in the gel filtration chromatogram: first peak at ~89 mL elution volume followed by a small second peak at ~99 mL elution volume. These peaks correspond to the molecular mass of the dimeric and monomeric Hpa2, respectively (**Figures 16A and 16B**). On increasing the concentration of the NaCl from 2 M to 4 M, the smaller peak representing the monomeric

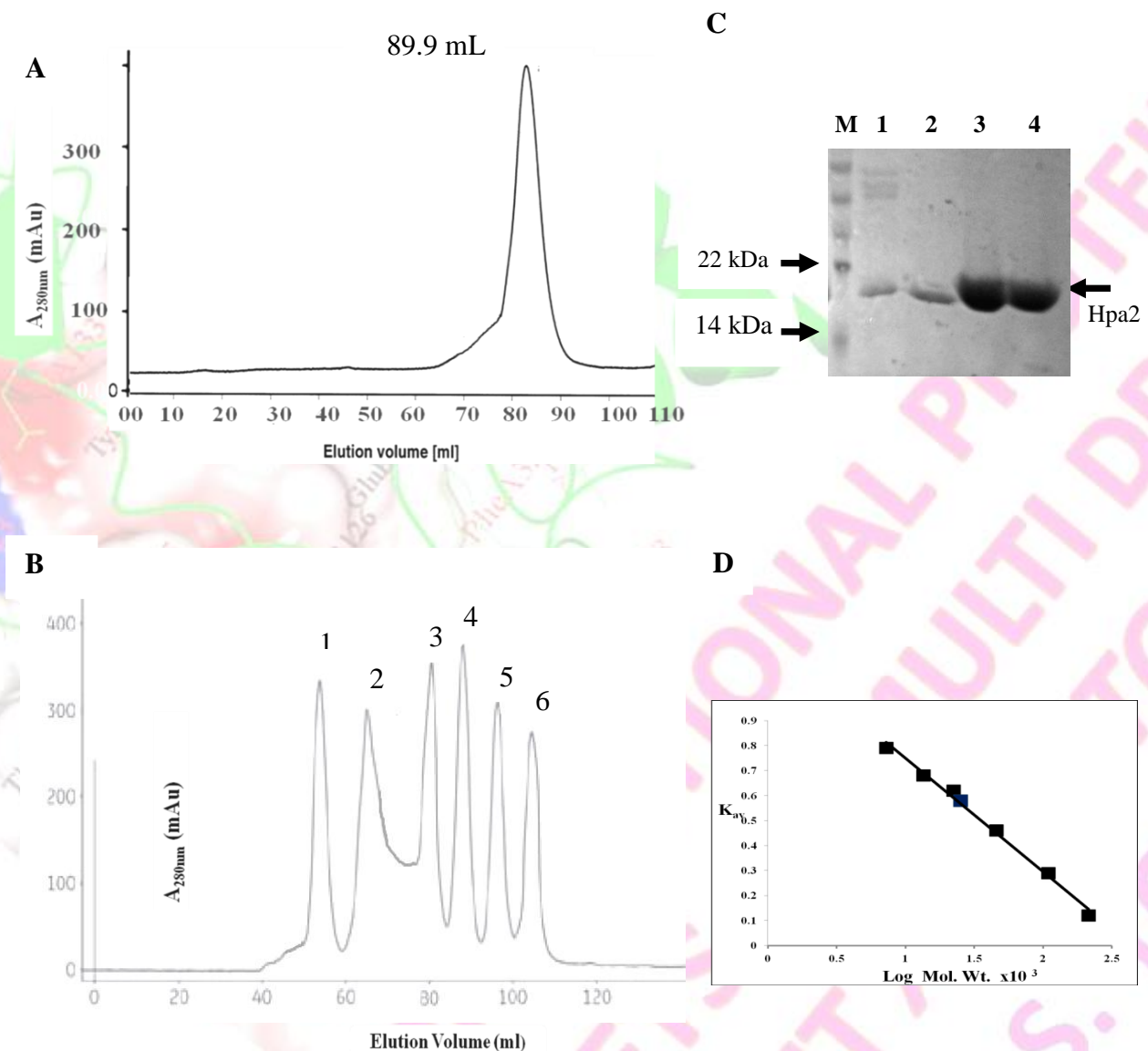


Figure 12. Chromatogram obtained after loading the Hpa2 protein on 16/600 Superdex 200 column. **(A)** The void volume of the column is 40 mL and after injection the protein peak appeared at around 89.6 mL of volume. **(B)** Chromatogram for standard proteins runs on 16/600 Superdex 200 column with their molecular weights and elution volume. 1) thyroglobulin [670 kDa, 49.92 mL], 2) aldolase [158 kDa, 62.86 mL] 3) ovalbumin [44 kDa, 74.11 mL], 4) carbonic anhydrase [29 kDa, 89.9 mL], (5) ribonuclease A [13.7 kDa, 94.56 mL], and 6) aprotinin [6.5 kDa, 104.10 mL]). **(C)** Protein fractions from gel filtration chromatography; M is the protein ladder, ~15 kDa Hpa2 is marked by an arrow. **(D)** Graph of log MW (kDa) vs. K_{av} for five standard proteins and Hpa2.

form of Hpa2 becomes more distinct at 4 M as compared to the peak during 2 M NaCl, demonstrating that higher concentration of the NaCl might partially disrupt the dimeric structure. It is known that high salt concentration compresses the solvation shell due to which the salt gets opportunity to interact with the protein surface charge to disrupt favorable ionic interactions by stabilizing the ion pairs. Here, partial destabilization of dimeric structure at high salt reflects the role of intermolecular weak polar interactions in the Hpa2 dimerization. Our results and other experimental evidences [40, 41] have revealed that a lower salt concentration facilitates the dimer formation, while high concentration of salt dissociates the dimer, in other words, favours the formation of monomers. Furthermore, in many studies [42] a conformational change is noticed with increased ionic strength. This conformational change would allow the monomer to rotate freely *i.e.*, contributes to the positive entropy.

Here, chaotropic agent urea is used to examine the stability and nature of interactions present at the interface. Analysis of elution profile indicated that alike NaCl, the 2 M urea initiated the dissociation of dimeric Hpa2 (Figure 13C). Conversion of dimeric Hpa2 into monomeric form was found to be more prominent in 4 M of urea than in NaCl (Figure 13D). The mechanism of denaturation process mediated by urea is a historical puzzle [43–45]. According to one of the well known mechanisms, denaturation by urea is a two-stage process, in the first stage, urea molecules accumulate in the vicinity of the protein surface and replaces the solvent molecules and in this way it initiates the unfolding of the protein. This means urea has better solvation capacity than water. In the second step, urea penetrates into the protein interior and forms hydrogen bonds with backbone atoms and disrupts the hydrophobic interaction. So the destabilizing nature of urea was endorsed to its direct H-bonding property. However, electrostatic interaction and van der Waals dispersion force is also considered as the driving force for the urea–protein interaction. The examination of gel-filtration chromatogram showing the elution profile of Hpa2 in 4 M urea designates that the second peak corresponding to the monomeric Hpa2, becomes more distinct and intense than 4 M NaCl. The higher dissociation in 4 M urea concentration indicates that inter-molecular hydrophobic interactions are major stabilizing force at the dimeric interface, which corroborates our *in silico* results.

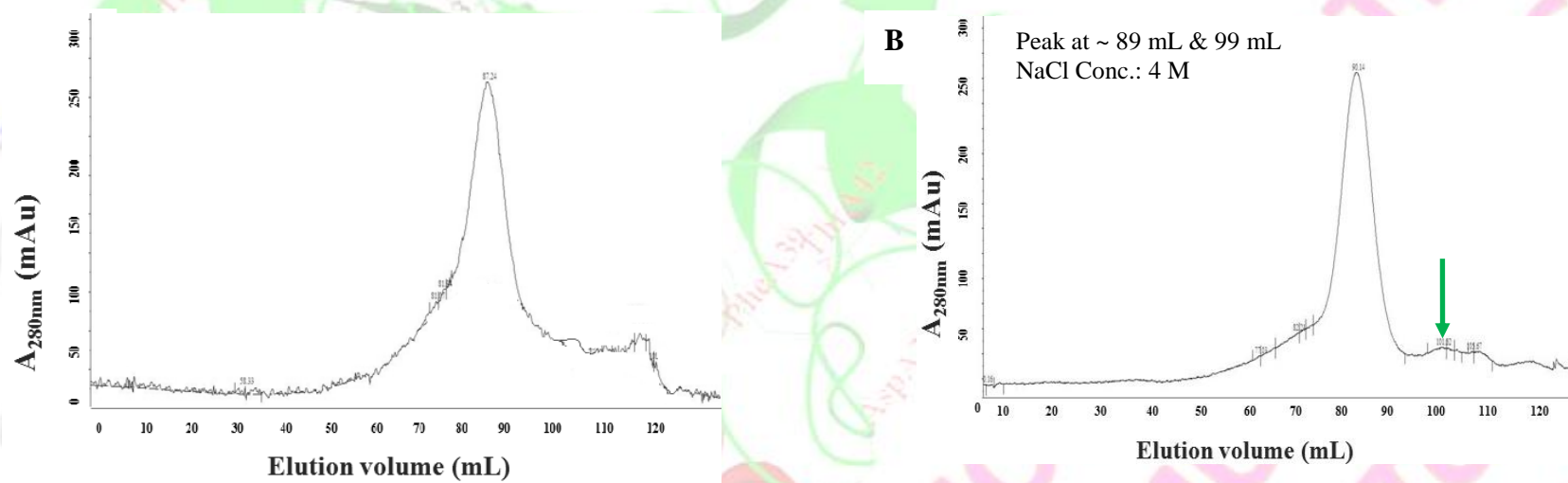


Figure 13. Effect of chaotropic agents NaCl and urea on oligomeric property of Hpa2. (A) Gel-filtration chromatogram showing the elution profile of Hpa2 protein treated with (A) 2 M of NaCl (B) 4 M NaCl (C) 2 M urea (D) 4 M urea.

3.5.5.2. Effect of substrates/inhibitors binding on the Hpa2 oligomeric state

Proteins homologous to Hpa2 are found to exist as monomeric (1q2y, 3efa, 2jdc, 2bsw), dimeric (1ghe, 1j4j, 2dxq and 1qsm), tetrameric (1qsm) and higher oligomeric (1xeb) states. As discussed in the *Introduction section* (3.1), the oligomeric state of Hpa2 depends on the presence of substrate. Furthermore, in biochemical study, Sternglanz and co-workers [46] noted that functional behaviour of protein Hpa2 is different from other acetyltransferases [46]. As it has capacity to acetylate polyamines – spermine, spermidine, and putrescine *in vitro*. In our *in silico* studies (*section 2.3.6*) we have found that Hpa2 have high binding affinity with the polyamines and binding strength is highest for spermine followed by spermidine and putrescine.

For studying the effect of acetyl-CoA on Hpa2, four different concentrations of acetyl-CoA (0.15 mM, 0.24 mM, 0.35 mM and 0.45 mM) were incubated with the Hpa2 (5 mg/mL) in 500 μ L reaction for 4–6 h at 4 °C and then uploaded in Superdex 200 column and after completion of run, elution profiles were analyzed, and thus obtained chromatograms are depicted in the **Figures 14A–D**. Examination of elution peaks indicates that in the presence of acetyl-CoA, the dimeric Hpa2 is converted into the monomeric Hpa2, in case of yeast, Hpa2 exists as stable dimer in solution forming tetramer in the presence of acetyl-CoA it forms [4] tetramer. However, in case of Hpa2 (*A. baumannii*) reverse oligomeric state is obtained.

At sufficient substrate concentration, approximately all dimeric molecules were converted into the monomeric form. From above results it is believed that in case of prokaryotes, monomeric form of Hpa2 is active unlike the eukaryotic cell in which it is active in tetrameric form. In Hpa2 each monomer unit have an individual active site and its oligomerization in case of eukaryotes might be related to increase in the number of binding sites.

As discussed in *Sections 3.1 and 3.4.5*, the dimeric structure of Hpa2 is very stable, consisting of an intertwined interface usually made up of hydrophobic residues. Stability of Hpa2 in high concentration of salt and denaturant was examined. It is noticed that dimeric structure remains almost intact even in extreme conditions. However, small concentration of the acetyl-CoA can convert Hpa2 into monomeric form. The conversion of Hpa2 into monomeric form is believed to be induced by the binding of acetyl-CoA *i.e.*, binding of acetyl-CoA might induce some conformational change in the dimeric form which leads its conversion into monomeric form, inverse to yeast Hpa2 where it changes to tetrameric form.

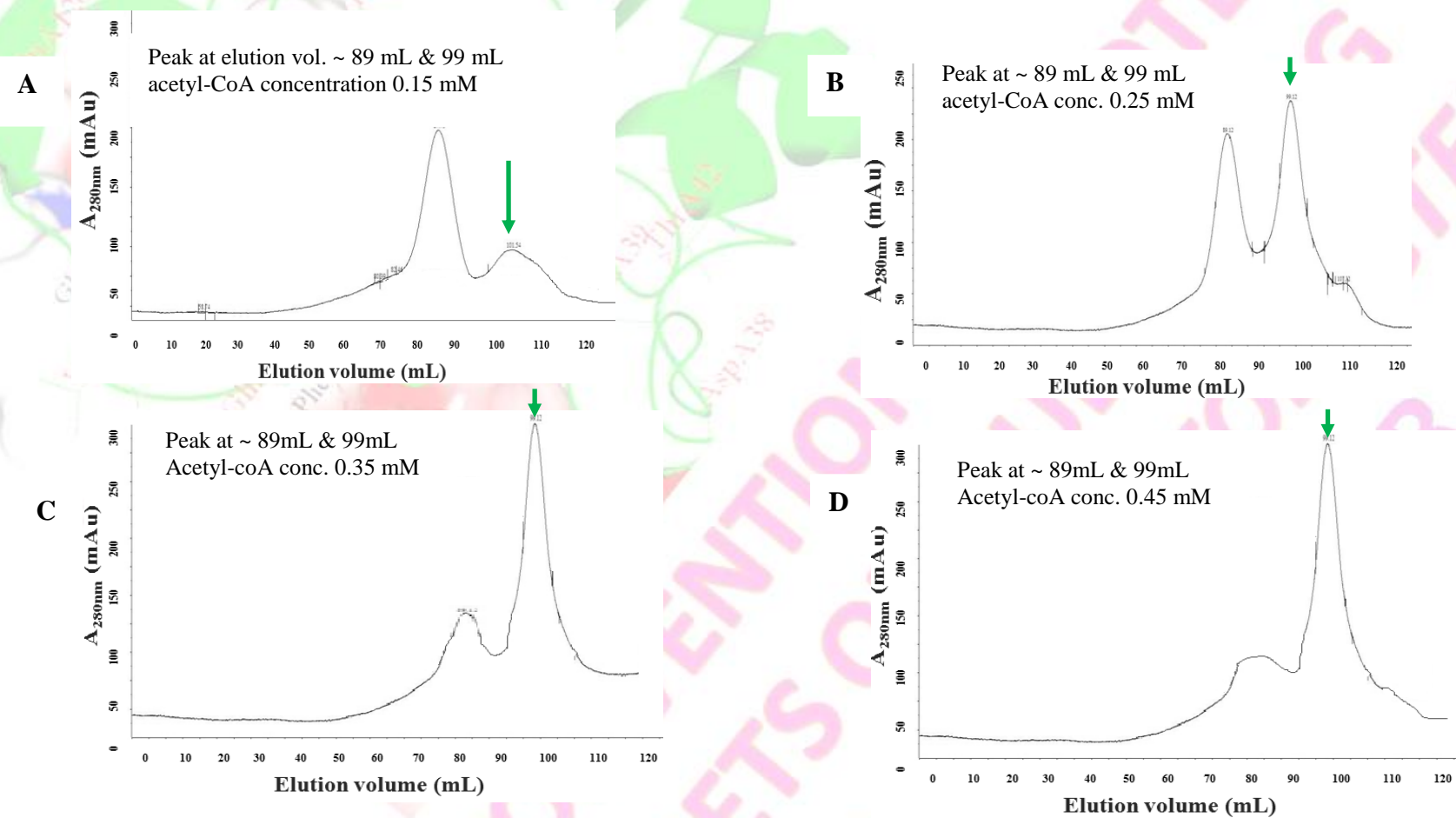


Figure 14. Effect of substrate (acetyl-CoA) concentration on oligomeric state of Hpa2. Gel filtration chromatogram showing the elution profile of Hpa2 (5 mg/mL) treated with acetyl-CoA of (A) 0.15 mM (B) 0.25 mM (C) 0.35 mM (D) 0.45 mM.

As explained in *Section 3.4.5.2* about the polyamine binding of Hpa2, here polyamine binding experiments were carried out in 500 μL reaction mixture. Each reaction mixture contains 3 mM concentration of polyamines and Hpa2. Analysis of size exclusion chromatogram infers that binding of polyamine also induces conformational change in the dimeric state of Hpa2. Interestingly, the binding of spermidine and spermine with Hpa2 protein resulted in the formation of higher oligomeric form *i.e.*, tetramer (**Figures 14A and 14B**). The change in the oligomeric state might be related to higher acetylating capacity of Hpa2 and binding of spermidine / spermine are supposed to induce the conformational changes in Hpa2. Putrescine binding does not affect at all and its oligomeric state corroborating its weak binding strength (**Figure 14C**).

The polyamines acetylation is considered to be essential for their breakdown as well as for their export from the cells. It was observed by Sternglanz and co-workers [46], that unlike the enzyme Paa1, the overexpression of yeast Hpa2 did not exhibit any growth defects on the culture media lacking the pantothenate, and *Hpa2* mutant does not show additional sensitivity to hydroxyurea like a *paa1* mutant; these two aspects exhibited by *Hpa2* suggest that polyamine acetylation might not be the primary role associated with Hpa2 in the cells. The interface of yeast Hpa2 crafted by hydrophobic residues and sequence alignment showed (*Section 2.4.2*) similarity in the interface residues indicate that *A. baumannii* recombinant Hpa2 also have hydrophobic interface having hydrophobic interactions as the main dimer stabilizing forces. Stability that induces strong interaction is not only contributed by the interactions between side chains of hydrophobic amino acids but also by the large increase in entropy due to the removal of hydrophobic surface area from ordered solvating water.

It is known from the previous studies that about $1.5 \text{ kcal mol}^{-1}$ of favourable energy are contributed by each methylene group which is coming by removing ordered solvation water. It is widely accepted that urea breaks the structured-water to induce the denaturation, and in presence of large amount of urea, the effective driving force (a subtle balance between hydrophobic interactions and interfacial free energies) for compact structure formation in proteins decreases leading to a destabilization of the native state. Similarly, in this study, dimer structure is found to be disrupted more by urea than by NaCl attesting to statements that interface is hydrophobic in nature. This interpretation is also in line with our *in silico* results discussed in *Chapter 2*. Further, oligomerization is induced by the polyamines (spermine, spermidine) not by acetyl-CoA. In the *in silico* work, we found that irrespective of sequence

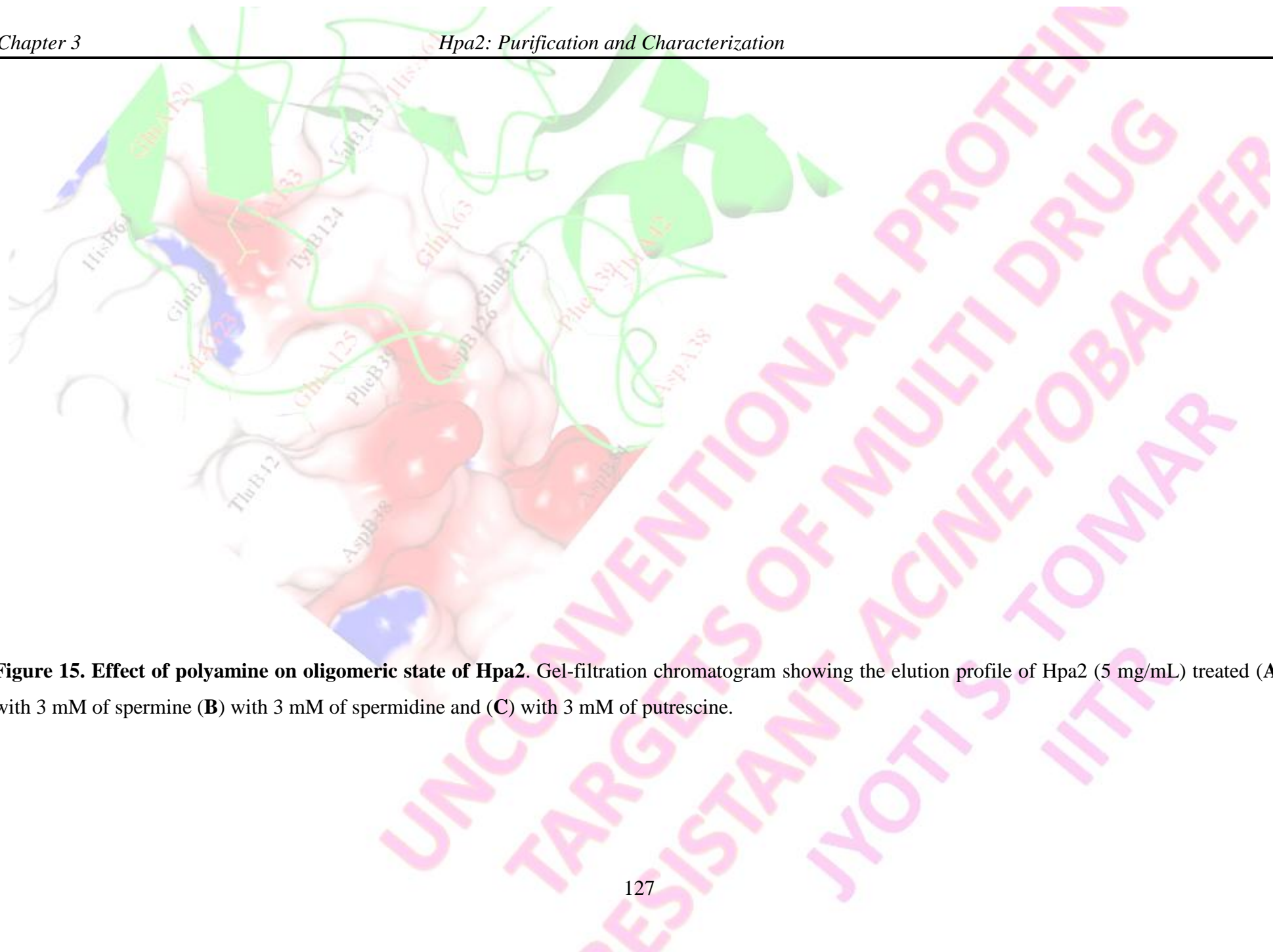


Figure 15. Effect of polyamine on oligomeric state of Hpa2. Gel-filtration chromatogram showing the elution profile of Hpa2 (5 mg/mL) treated (A) with 3 mM of spermine (B) with 3 mM of spermidine and (C) with 3 mM of putrescine.

similarity, the dimeric interface of *Acinetobacter baumannii* Hpa2 is more similar to the yeast Hpa2 interface, but not to its homologue 1xeb. In 1xeb, two types of dimers are formed in which one mainly consists of the residues from the secondary structure $\beta 1$, $\alpha 1$, $\alpha 2$ and $\beta 2$ (A–D) which showed the rmsd value (~ 10.5). Second one is A–C dimer which involves the secondary structures $\beta 3/\beta 4$, $\beta 4$, $\beta 5$, $\alpha 3$, $\beta 6/7$ and $\beta 8$. From the sequence similarity and rmsd value (~ 4.5), it is inferred that second (A–C) dimeric structure is more considerable in our case.

3.5.6. Native PAGE

Profile of Hpa2 at different molar concentrations was analyzed by native PAGE as described in the *Methodology section* 3.4.9 [47]. The pure protein, without any further treatment, was added with gel loading buffer, heated for 30 sec at 90 °C and then loaded on the 12% gel. Pure and single band of approximately ~ 30 kDa was observed on the gel (Figure 16A) indicating towards the existence of Hpa2 as dimeric state in solution, which is independent of protein concentration. The native gel of Hpa2 in presence of various substrates is depicted in Figure 16B. The Hpa2 exists as dimer in solution converting completely into monomeric form on binding with acetyl-CoA substrate (Lane: L2). Hpa2 gets converted into higher oligomeric state *i.e.*, tetrameric form in the presence of spermine and spermidine. The progress of two dimers towards higher oligomeric state is clearly visible in gel (Lanes: L3 and L4). Dimeric structure of Hpa2 although remains intact upon binding of putrescine (Lane: L5).

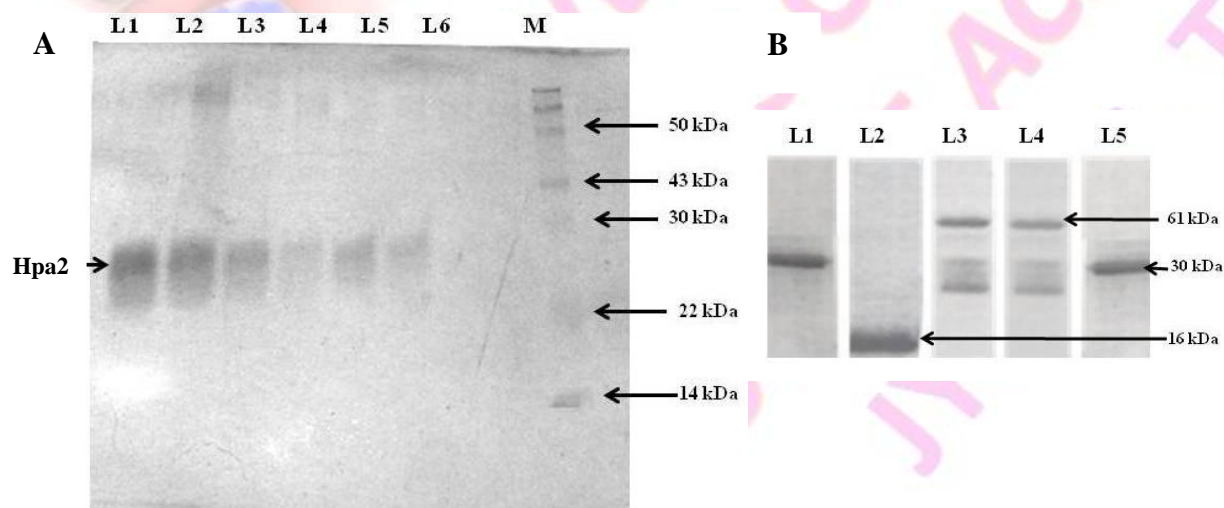


Figure 16. Native PAGE gel for Hpa2. (A) Lanes L1–L6 : Hpa2 concentration (3–0.5 mM) Lane M : prestained protein ladder. (B) Lane L1 : Hpa2 alone, Lane L2 : Hpa2 + acetyl-CoA, L3 : Hpa2 + spermine, L4 : Hpa2 + spermidine, L5: Hpa2 + putrescine.

Native PAGE results and analytical results from size exclusion chromatography support the notion that Hpa2 oligomeric state modifies in presence of different substrates; however, it is independent of the concentration.

3.5.7. Analysis of Hpa2 secondary structure

CD was used for determination of secondary structure and folding of recombinant Hpa2 protein. Hpa2 exists as dimer (homodimer) with interface primarily made up of hydrophobic residues. It is known from literature that hydrophobic interaction becomes more stable with increase in the temperature. To corroborate the nature of interactions at dimeric interface as well as the stability of dimer, we performed CD titration by varying the temperature (20–65 °C). The far-UV CD spectrum obtained at 20 °C (190–260 nm) was analyzed by Dichroweb program using CDSSTR Method [48–50]. Analysis of spectra using the CDSSTR method showed ~ 42% content of the α -helix and 37% content of the β -sheets at physiological pH (pH 7.6) (Figure 17). These studies confirm that recombinant Hpa2 protein maintains its integrity during the purification process.

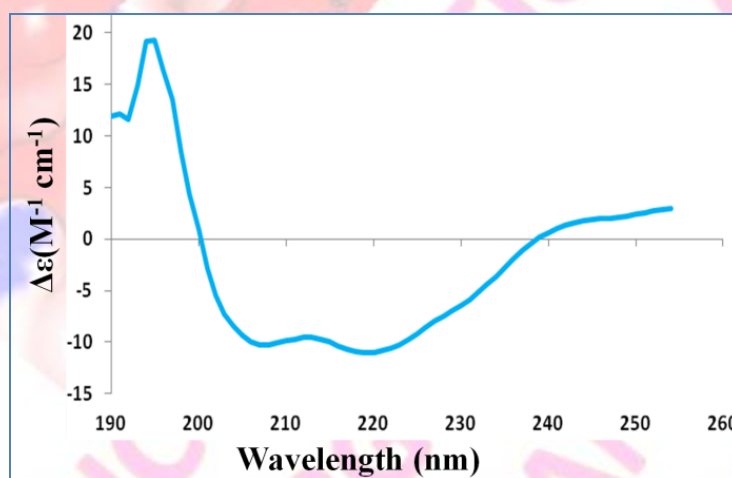


Figure 17. CD spectrum of protein Hpa2 0.25 mg/mL (50 mM phosphate buffer) in far-UV at pH 7.6 at 20 °C.

Generally, thermal denaturation experiment provides information about the stability of the folding process (*i.e.*, whether it is reversible or irreversible), T_m , unfolding intermediates and free energy of unfolding, *etc.* Here, thermal denaturation was carried out and the outcome of the spectra was examined and plotted (Figure 18).

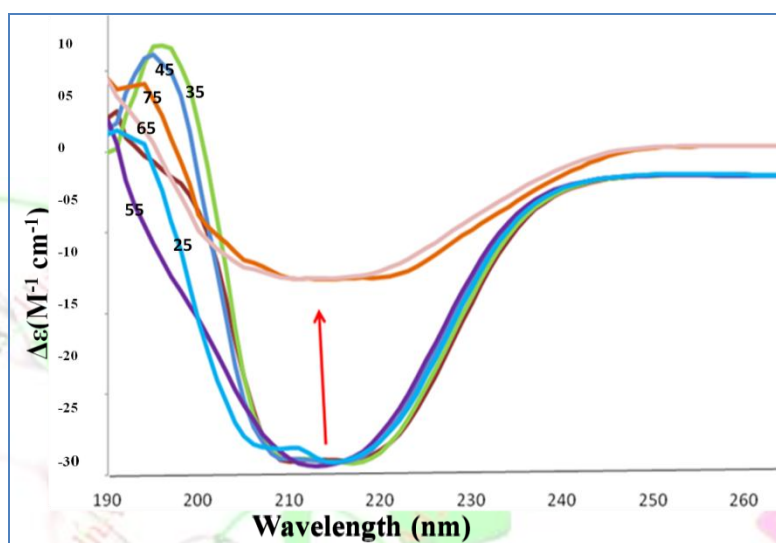


Figure 18. CD spectra of Hpa2 (1 mg/mL) during the course of thermal denaturation from 20 °C temperature in 50 mM phosphate buffer at pH 7.6. The temperatures of the experiment are shown along the spectra (25–75 °C).

The analysis of spectra recorded at different temperatures indicates that initial temperature increment resulted in nominal increase in secondary structures content. The secondary structure becomes more compact and organized up to 35 °C of temperature. However, as temperature goes beyond 45 °C, integrity of secondary structure is observed to be affected. As the temperature crosses range of 55 °C, considerable decrease in the CD spectra was found **which reflects the** changes in the native structure. From the above *Sections* (3.4.5, 3.4.9 and 3.4.10) studies it is known that Hpa2 exists as dimer in the solution. Absence of considerable changes in the CD spectra till 55 °C temperature designates that the dimeric structure is stable and remains intact up to this temperature. Initial increment in CD spectra with temperature (25–45 °C) could be the outcome of increase in the strength of interface hydrophobic interactions and stability of the dimer ($T_m = 55$ °C).

We also analysed the effect of heating on the secondary structures by measuring the percentage of helix. The scan was performed with the scan rate 1 °C per 5 min and secondary structure was calculated. We found that helix remains almost intact up to 45 °C and after this, opening of helix begins. At the temperature 55 °C, the opening of helix is rapid and more than almost 80% of helical structure opens at the temperature 75 °C. After heating of the sample up to 75 °C we started the cooling process to see whether the Hpa2 helix has the capacity to fold back. During cooling, no (5%) recovery in the helicity was noticed. This indicates that the

thermal denaturation of Hpa2 is an irreversible process, as cooling process does not refold the protein back to its native state.

Change in the CD signal at the 222 nm was plotted as a function of temperature (Figure 19), and a sigmoid denaturation curve was obtained, indicating that the denaturation of Hpa2 in the solution is a two-state process. Dramatic decrease in CD signal at 222 nm at higher temperature was also found. The increase of the random coil by high temperature is resulted in a dramatic decrease in CD signal at 205–220 nm (helix and sheet). Analysis of denaturation curve demonstrates that unfolding is accompanied by the formation of thermodynamically stable intermediate having intact secondary structure, no tertiary structure (data from near-UV CD spectra) was observed for Hpa2. Therefore, they may be classified as molten globule states. Far-UV CD spectra of the intermediates are mixture of α and β secondary structures conformations, unlike the native state, which is relatively rich in α -structures (Figures 17). It is examined through the studies of conformational transitions on numerous proteins at high temperatures which have shown that different proteins may undergo transitions through different intermediate states depending on the protein architecture and nature of denaturants.

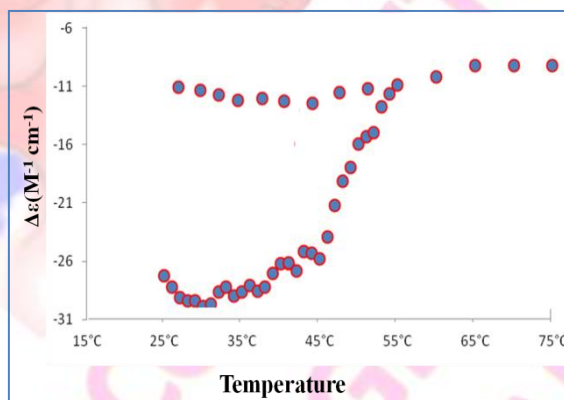


Figure 19. Far-UV CD spectra showing the effect of temperature on the Hpa2 stability, thermal denaturation of protein was monitored by CD at 222 nm, protein conc. 1 mg/mL in 50 mM phosphate buffer at pH 7.6, for the melting experiment the protein was heated from 20–75 °C.

3.5.8. Fluorescence spectroscopy

Fluorescence is a valuable technique to study the binding of ligands with protein to remark the conformational changes in protein on varying the physical or chemical environment [51–53].

Change in the microenvironment provides a higher sensitivity to the fluorophore which is the result of interactions between the fluorophore and adjacent groups [54–55]. These effects can be observed as a shift in the fluorescence maximum or as an increment/decrement in the fluorescence emission spectra. The effect of temperature and guanidinium hydrochloride (GdnHCl) on the Hpa2 fluorescence was studied (Figure 20A and 21). Fluorescence of the aqueous tryptophan was recorded between 290–450 nm. The denaturation curve is plotted as a function of fluorescence intensity ratio F_{350}/F_{330} vs. temperature (Figure 20B) as intrinsic fluorescence intensity ratio F_{350}/F_{330} is widely used as alternative probe for the denaturation studies.

Hpa2 protein contains two tryptophan residues which are partially buried in the nonpolar interior of proteins, due to which they give a fluorescence emission maximum in the range of ~330 nm. It is known that tryptophan residues, when come in contact with polar interfaces, exhibit fluorescence emission maximum ~350 nm. In thermal denaturation studies of Hpa2 from 20–45 °C temperature, fluorescence maxima were observed at 330 nm. On further heating a large red shift of 20 nm was noticed. This indicates conformational change in the structure of Hpa2 which is similar to the results from Hpa2 CD denaturation studies.

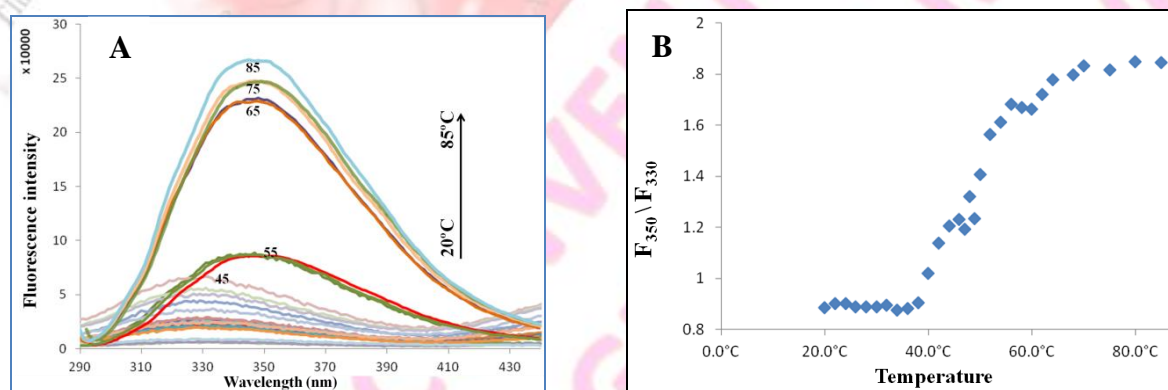


Figure 20. (A) Emission spectra of aqueous tryptophan at various temperatures. (B) The ratio of the intensities at 350 nm and 330 nm (λ_{\max}) at various temperatures.

Furthermore, like CD denaturation, here also above 55 °C temperature, dramatic increase in emission intensity is observed. Denaturation through GdnHCl is also informative about the folding-unfolding of Hpa2. Variation in the pattern of fluorescence spectra through thermal and GdnHCl, denaturation indicated that hydrophobic interactions have more impact than electrostatic interactions on the stability of the protein (as shown in Figure 20). In

addition to the increase in fluorescence intensity, the rise temperature also leads to a bathochromic shift or red shift of the fluorescence maximum from 335 nm (native form) to 350 nm (denatured form). This is due to the fact that the Trp residues, which are partly buried in the native form, are exposed on the surface in the denatured protein. As a consequence, tryptophan residues become hydrated upon the loss of structural integrity. Concomitantly, the intensity increases approximately by a factor of 3.0 (Figure 20A). As we have seen in CD spectra, here also the denaturation starts after 45 °C and red shift is noticed after this range, which could be the result of opening of dimeric structure. Above 55 °C, a large increase in the fluorescence intensity is observed which shows the fast denaturation of the protein. The thermogram exhibits a shift between 55 °C and 65 °C, indicating a second transition upon thermal unfolding of the Hpa2. It is inferred that above 55 °C, the tertiary structure is collapsed in a sharp two-state manner (Figure 20B) and the denaturation is completed around 75 °C. The ratio of the intensities at λ_{\max} 330 nm and 350 nm is calculated and plotted against temperature (Figure 20B). Similar to CD, the denaturation curve obtained is sigmoid and designates that denaturation of the Hpa2 is a two-state process.

The effect of the GdnHCl on the recombinant Hap2 is also studied under isothermal conditions. The unfolding/folding transitions were monitored by the changes in the fluorescence intensity and the position of the maximum of the Trp36 band at 20 °C. As shown in Figure 20, the increase in fluorescence intensity and bathochromic shift of the fluorescence maximum is observed after 2 M. These results indicate that dimeric form of Hpa2 remains intact till 2 M GdnHCl.

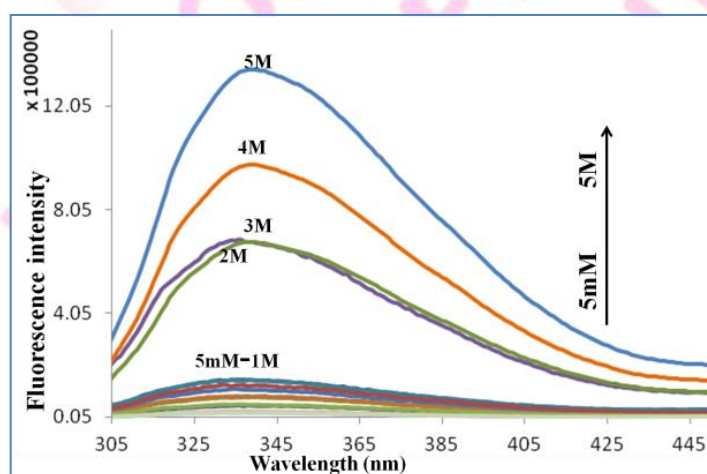


Figure 21. Emission spectra of aqueous tryptophan at various urea concentrations.

3.5.9. Isothermal titration calorimetry binding experiment of Hpa2 with substrates

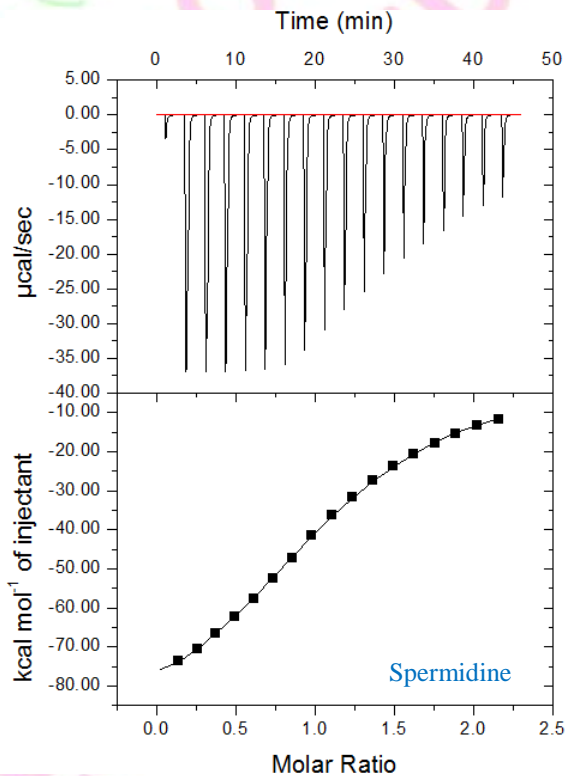
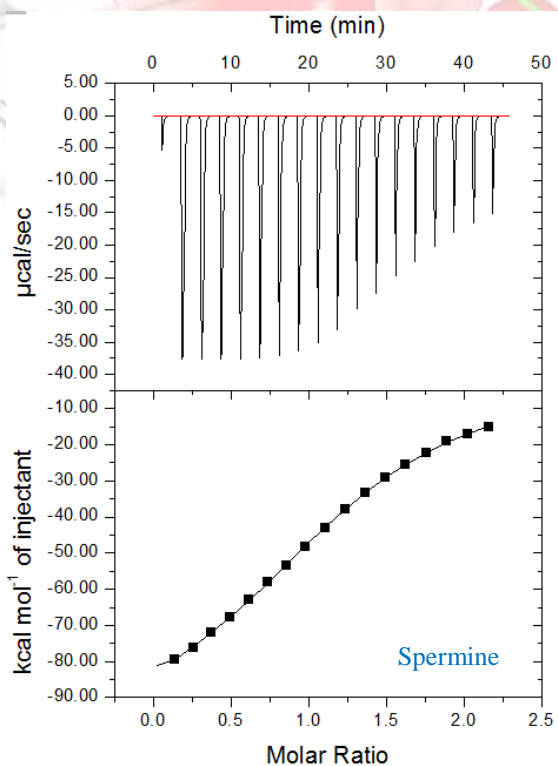
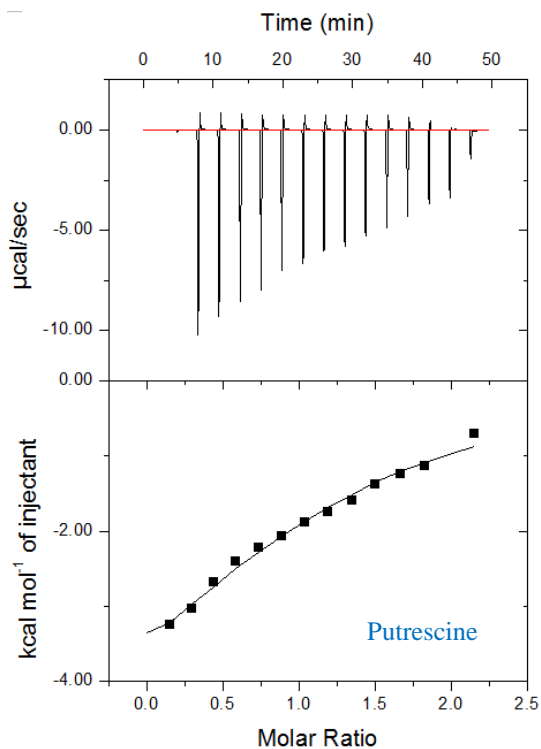
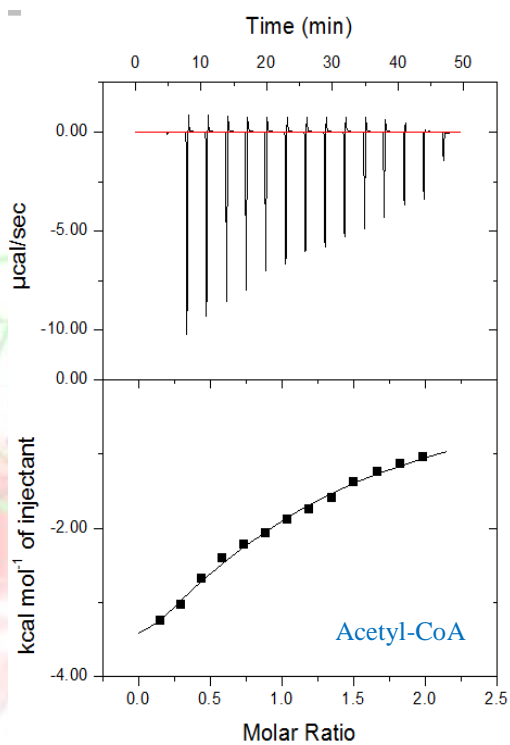
ITC is used in the field of protein science because of its wide application. It is easy to perform, and in the results, large amount of thermodynamic data was generated using small amount of protein [56–59]. Secondly, the determination of parameters K_b , N , ΔH and ΔS in a single experiment is possible. Experiments are performed with significant ease and also large amount of thermodynamic data is gathered using small quantity of the protein through ITC studies [60, 61]. Parameters like K_b , N , ΔH and ΔS can be determined through single experiment in case of ITC. Due to these reasons, ITC found wide application in the field of protein science. Till date, no thermodynamic data is available in literature for Hpa2 binding with its substrates polyamines and acetyl-CoA, which guided us to perform these experiments. Here, titration of Hpa2 with acetyl-CoA, polyamines and antibiotics was performed to analyze the binding parameters and substrate preference of Hpa2.

Analysis of binding isotherm signifies that Hpa2 has strong binding with the substrates with a differential preference, depicted in Figure 21. Binding parameters indicate that binding of acetyl-CoA to Hpa2 was driven by a favourable enthalpy change ($\Delta H = -8.903 \pm 0.2024$ kcal mol⁻¹) in accordance to our *in silico* prediction. This further confirms that the binding of acetyl-CoA to the active site of enzyme is thermodynamically favorable. One molecule of Hpa2 binds with one molecule of acetyl-CoA ($N = 0.937 \pm 0.272$) which is similar to our size exclusion chromatography results (during determination of oligomeric state in the presence of substrate showing binding of acetyl-CoA to the monomer Hpa2). The binding constants K_a and K_d are $(3.75 \pm 0.14) 10^2$ M⁻¹ and 266 ± 0.02 μ M, respectively. The change in entropy is -12.2 cal mol⁻¹ deg⁻¹, the Gibbs free energy change of binding is calculated using the equation; $\Delta G = \Delta H - T\Delta S$, was found to be -5.27 kcal mol⁻¹. The calculated binding parameters values, are similar to binding parameters of acetyl-CoA with the Pat protein which also belongs to the GNAT family [62], attesting to the values.

On the basis of *in silico* studies of Hpa2 (Section 2.3.6), binding studies of Hpa2 with the antibiotics was also performed. Thermodynamic parameters of Hpa2 and its interactions with acetyl-CoA, polyamines and antibiotics are presented in Table 1. Analysis of this data clearly indicates that the binding of polyamines are different from the substrate acetyl-CoA, which has only one binding site while polyamines possess more than one binding site. The binding affinity of spermine is found to be highest among the all polyamines followed by

spermidine and putrescine, which has lowest binding affinity. These results are in agreement with the results obtained by Sternglanz and co-workers during the biochemical characterization of yeast Hpa2 [46]. They also concluded that Hpa2 are able to acetylate all three polyamines with a preference for spermine and spermidine over putrescine [46]. Furthermore, the acquired ITC parameters depicting binding strength attest the results obtained from *in silico* studies (Section 2.3.6). The N values of polyamines are ~ 1.6 which indicates that polyamines possess two binding sites and size exclusion chromatography result also describes Hpa2 to exist as tetramer having two binding sites like yeast Hpa2. As already discussed, yeast Hpa2 forms tetrameric structure in presence of acetyl-CoA and this tetrameric structure contains two binding sites [4]. The polyamines binding with Hpa2 occurs due to favourable enthalpy change ($\Delta H < 0$) while entropy change ($T\Delta S < 0$) hinders the binding process. This is similar to the interactions of acetyl-CoA with the aminoglycoside acetyltransferase enzyme, an enzyme which possesses almost similar secondary and tertiary structures as Hpa2 and is also a member of the same superfamily [62].

Binding of spermine and spermidine does not exhibit remarkable difference in change in Gibbs free energy value; however, changes in enthalpy ($-\Delta H \sim 100$ kcal/mol) and entropy ($+T\Delta S = \sim 95$ kcal/mol) values are high. This might be due to the association of Hpa2 dimer to form the tetramer achieved through formation of many interactions at the interface, and these interactions favour the enthalpy change [63]. Tetrameric structure is known to be formed by polar residue interactions [4] enthalpy change for separating polar amino acids from the water is negative and entropy change for the packing of amino acid is negative because the folded state is more ordered than the unfolded one attesting to the high values of ΔH and ΔS . In the case of putrescine, binding is not so significant. Similar to size exclusion chromatography results in which the elution profile confirms that no change in the oligomeric state takes place on putrescine binding. Similarly, ITC results also indicate that putrescine is unable to induce the conformational changes. It does not exhibit strong binding strength ($\Delta G = -3.2$ kcal/mol) as depicted by the spermine ($\Delta G = -5.2$ kcal/mol) and spermidine (-4.5 kcal/mol); however, binding nature is same for all the polyamines *i.e.*, enthalpy driven. Binding of polyamines is not favored in terms of change in entropy as complex formation decreases the degree of freedom.



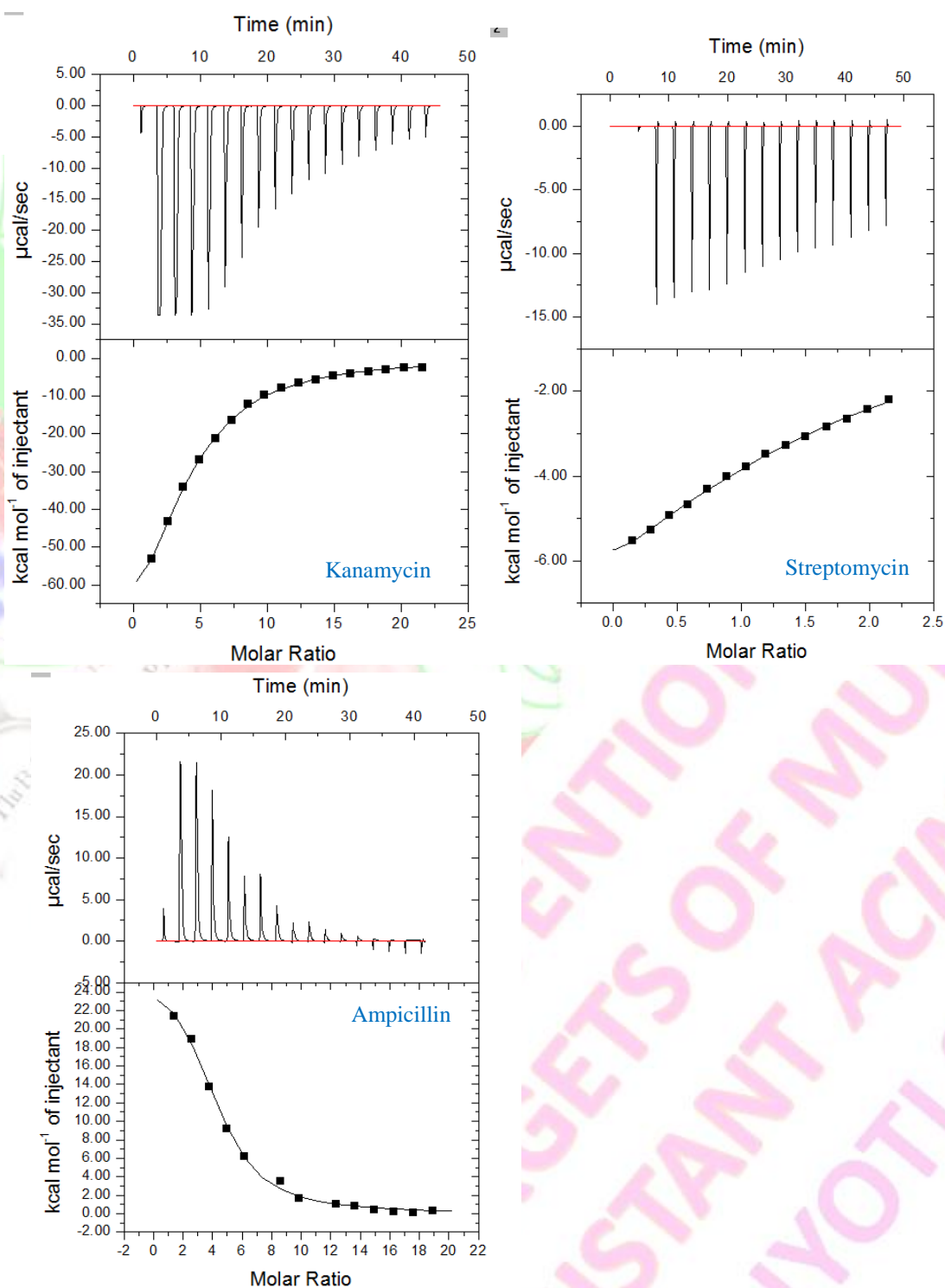


Figure 22. ITC profile of acetyl-coA, polyamines and antibiotics binding to Hpa2. Raw data from titration of consecutive injections of substrates into Hpa2, represented as the heat change ($\mu\text{cal/s}$) upon injection over time (Top), binding isotherm obtained by integration of the raw data (reported as kcal/mol). The solid line represents the best-fit curve generated from a one-site binding model (Bottom).

Thermodynamic parameters for binding polyamines and antibiotics with the Hpa2					
Substrate	N	K_D (μM)	ΔH ($\text{kcal}\cdot\text{mol}^{-1}$)	$-T\Delta S$ ($\text{kcal}\cdot\text{mol}^{-1}$)	ΔG ($\text{kcal}\cdot\text{mol}^{-1}$)
Acetyl-CoA and Polyamines					
Acetyl-CoA	0.94 ± 0.27	266 ± 0.02	-8.903 ± 0.20	3.63 ± 0.17	-5.27
Spermine	1.63 ± 0.04	160.51 ± 0.11	-102.9 ± 0.62	97.74 ± 0.17	-5.16
Spermidine	1.61 ± 0.05	143.47 ± 0.13	-95.53 ± 0.64	90.99 ± 0.13	-4.51
Putrescine	1.18 ± 0.12	91.70 ± 0.30	-5.95 ± 0.11	01.81 ± 0.11	-3.25
Aminoglycoside antibiotic					
Ampicillin	3.99 ± 0.12	75.75 ± 0.17	27.76 ± 0.11	-33.37 ± 0.29	-5.67
Streptomycin	1.68 ± 0.02	169.74 ± 0.8	-11.56 ± 0.83	7.78 ± 0.24	-3.78
Kanamycin	3.74 ± 0.08	259.06 ± 0.13	-101.50 ± 0.29	96.55 ± 0.38	-4.95

Table 1. Thermodynamic parameters for binding of polyamines and antibiotics to Hpa2. Given errors were calculated as the standard error of the mean of two trials.

ITC experiments allowed the calculation of the contributions that enthalpy (ΔH) and entropy (ΔS) changes have on binding of antibiotics to Hpa2. For binding of aminoglycoside class of antibiotics *i.e.*, kanamycin and streptomycin, most of the free energy of association was ($\Delta G = -4.95$ kcal/mol; -3.78 kcal/mol) contributed by the enthalpy terms ($\Delta H = -101.5$ kcal/mol; -11.56 kcal/mol, respectively). But entropy did not favor this binding (as in the case of polyamines) with Hpa2, thus it is driven by the enthalpy term. The similarity in nature of binding indicates that these antibiotics possess similar binding sites. Kanamycin binding has been observed to be inducing oligomerization of Hpa2 similar to the case of spermine and spermidine. However, streptomycin binding isotherm matches with that of the putrescine and both of these were not found to induce any conformational changes in the protein structure. Unlike these aminoglycoside antibiotics, ampicillin possesses different binding property *i.e.*, entropy driven ($T\Delta S = -33.37$ kcal/mol). The entropy driven binding of ampicillin can be explained by a model according to which, the binding site is filled by the water molecules that are relocated upon ligand binding, concurrently producing a positive entropic contribution to ΔG . Additionally, desolvation of ampicillin, which may be a requirement for binding, could also increase ΔS . The analysis of binding data for streptomycin ($N = 1.68$) and kanamycin ($N = 3.74$) reveals significant binding affinity. The higher number of binding sites for kanamycin

and ampicillin are found ($N = 4$), which might be the result of non-specific binding or they might have capacity to induce higher oligomeric state in the Hpa2, *i.e.*, tetramer [64].

3.6. Conclusion

The Hpa2 protein of MDR bacteria *A. baumannii* was cloned in pET28a vector and heterologous expression of the protein was done in BL21 cells. The expressed protein was purified up to 98% homogeneity by a three-step purification protocol involving affinity purification followed by the cleavage of the 6xHis-tag using TEV protease and its removal using reverse affinity step. In the final step, size exclusion chromatography was used to separate the minor impurities present in the protein.

CD experiments were performed to determine the secondary structure and the analysis of results demonstrate that Hpa2 contains properly folded secondary and tertiary structures. The denaturation study was done by varying physical and chemical environment using CD and fluorescence experiments. Denaturation graphs plotted infer that Hpa2 is a stable protein and its denaturation is a two-step process. It is known from the literature that oligomeric conformation of Hpa2 depends on the substrates. So here oligomeric state of Hpa2 is determined in presence of various substrates as well as in extreme physiological conditions using analytical size exclusion, native PAGE and glutaraldehyde crosslinking assay. The results from these experiments attest that Hpa2 exists as dimer in the solution and oligomeric state is independent of its concentration. Hpa2 shows different oligomeric states in the presence of different substrates. Furthermore, calorimetric binding studies show that binding of substrates with Hpa2 is driven by favourable changes in enthalpy except for ampicillin. Based on the analysis of the results of experiments performed, it is concluded that Hpa2 might have capacity to acetylate various substrates like histone, polyamines and aminoglycoside antibiotics. The Hpa2 protein was put onto crystallization trials using the commercial Hampton screens *via* sitting drop vapour diffusion method. Initial hit of the Hpa2 crystals provided us the condition suitable for the crystallization. Further trails are required for obtaining protein crystals for the complete elucidation of structural and functional properties of Hpa2.

3.7. References

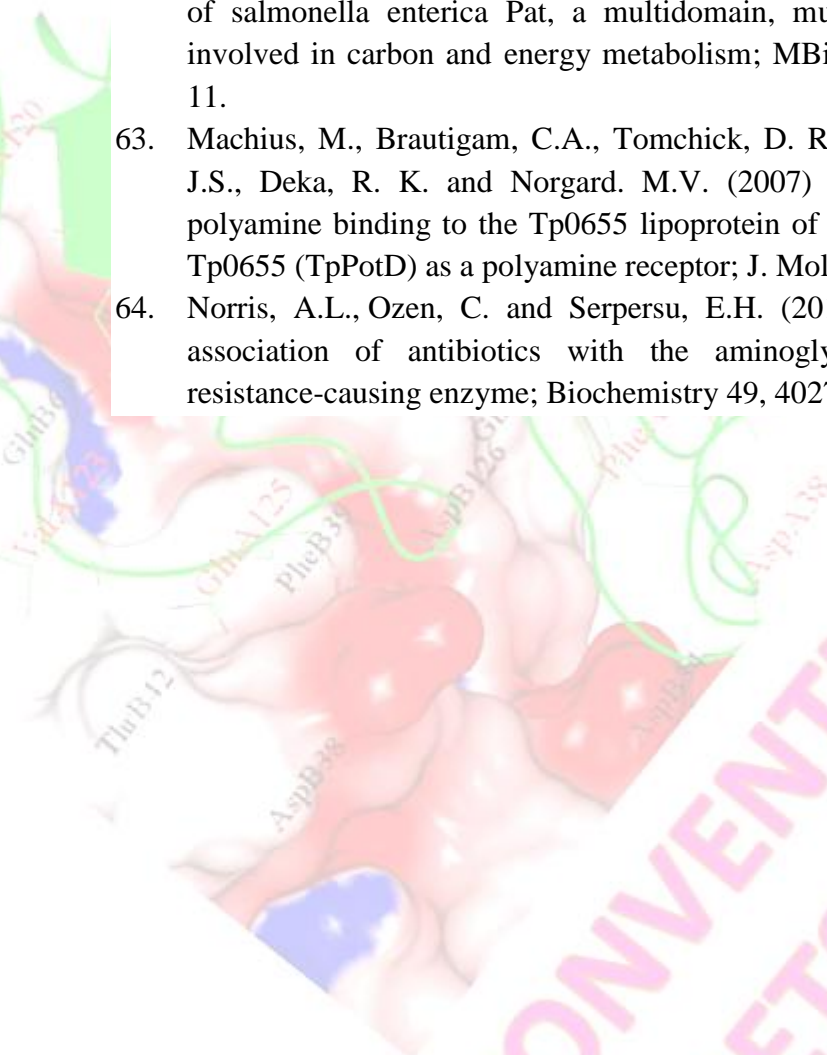
1. Schreckenberger, P.C., Daneshvar, M.I., Weyant, R.S. and Hollis, D.G. (2007) *Acinetobacter*, *Achromobacter*, *Chryseobacterium*, *Moraxella*, and other nonfermentative gram-negative rods; In: Murray, P.R., Baron, E.J., Jorgensen, J.H., Tenover, M.C. and Tenover, R.H. editors. Manual of clinical microbiology 8th ed. Washington, DC: American Society for Microbiology Press, pp. 770–779.
2. Fournier, P.E. and Richet, H. (2006) The epidemiology and control of *Acinetobacter baumannii* in health care facilities; Clin. Infect. Dis. 42, 692–699.
3. Ku, S.C., Hsueh, P.R., Yang, P.C. and Luh, K.T. (2000) Clinical and microbiological characteristics of bacteremia caused by *Acinetobacter lwoffii*; Eur. J. Clin. Microbiol. Infect. Dis. 19, 501–505.
4. Angus-Hill, M.L., Dutnall, R.N., Tafrov, S.T., Sternglanz, S. and Ramakrishnan, V. (1999) Crystal structure of the histone acetyltransferase Hpa2: A tetrameric member of the Gcn5-related *N*-acetyltransferase superfamily; J. Mol. Biol. 294, 1311–1325.
5. Protein Purification: Principles, High Resolution Methods, and Applications; (1998) Janson, J.-C. and Ryden, L. (Eds), Wiley-VCH.
6. Fish, W.W., Reynolds, J.A. and Tanford, C. (1970) Gel chromatography of proteins in denaturing solvents; J. Biol. Chem. 245, 5166–5168.
7. Pancoska, P., Blazek, M. and Keiderling, T.A. (1992) Relationship between secondary structure fractions for globular proteins. Neural network analysis of crystallographic data; Biochemistry 31, 10250–10257.
8. Compton, L.A., and Johnson, W.C., Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication; Anal. Biochem. 155, 155–167.
9. Woody, R.W. (1992) Circular dichroism and conformation of unordered peptides; Adv. Biophys. Chem. 2, 37–79.
10. Sreerama, N., Venyaminov, S.Y. and Woody, R.W. (2000) Estimation of protein secondary structure from CD spectra: Inclusion of denatured proteins with native proteins in the analysis; Anal. Biochem. 287, 243–251.
11. Yang, J.T., Wu, C.-S. C. and Martinez, H.M. (1986) Calculation of protein conformation from circular dichroism. Methods; Enzymol. 130, 208–269.
12. Roumita, M. and Jayant, B.U. (2014) Thermodynamic Characterization of the Unfolding of the Prion Protein; Biophys. J. 106, 410–420.
13. Goluguri, R.R. and Udgaonkar, B. (2015) Rise of the Helix from a Collapsed Globule during the Folding of Monellin; Biochemistry 54, 5356–5365.
14. Arndt, C., Koristka, S., Bartsch, H. and Bachmann, M. (2012) Native polyacrylamide gels; Methods Mol. Biol. 869, 49–53.
15. Fadoulglou, V.E., Kokkinidis, M. and Glykos, N.M. (2008) Determination of protein oligomerization state: two approaches based on glutaraldehyde crosslinking; Anal Biochem. 373, 404–406.

16. Bradford, M.M. (1976) Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding; *Anal. Biochem.* 72, 248–254.
17. Johnson, W.C. (1999) Analyzing protein circular dichroism spectra for accurate secondary structures; *Proteins: Struct. Funct. Genet.* 35, 307–312.
18. Johnson, W.C., Jr. (1988) Secondary structure of proteins through circular dichroism spectroscopy; *Annu. Rev. Biophys. Biophys. Chem.* 17, 145–166.
19. Giri, P. and Suresh, K.G. (2007) Specific binding and self-structure induction to poly (A) by the cytotoxic plant alkaloid sanguinarine; *Biochim. Biophys. Acta* 1770, 1419–1426.
20. Kruger, N.J. (1994) The Bradford method for protein quantitation; *Methods Mol Biol.* 32, 9–15.
21. Simanshu, D.K., Savithri, H.S. and Murthy, M.R.N. (2006) Crystal structures of *Salmonella typhimurium* biodegradative threonine deaminase and its complex with CMP provide structural insights into ligand-induced oligomerization and enzyme activation; *J. Biol. Chem.* 281, 39630–39641.
22. Fang, Y., Kolmakova-Partensky, L. and Miller, C. (2007) A bacterial arginine–agmatine exchange transporter involved in extreme acid resistance; *J. Biol. Chem.* 282 176–182.
23. Onica, G., Mota, D., Calugaru, A., Manciualea, M. and Dima, S. (1983) Immunogenicity and effector functions of glutaraldehyde-treated rabbit and mouse immunoglobulin, G; *Mol. Immunol.* 20, 709–718.
24. McIntosh, D.B. and Ross, D.C. (1985) Role of phospholipids and protein–protein associations in activation and stabilization of soluble Ca²⁺ -ATPase of sarcoplasmic reticulum; *Biochemistry* 24, 1244–1251.
25. Kapoor, M. and O'Brien, M.D. (1977) Investigation of the quaternary structure of *Neurospora* pyruvate kinase by cross-linking with bifunctional reagents: the effect of substrates and allosteric ligands; *Can. J. Biochem.* 55, 43–49.
26. Kiernan, J.A. (2000) Formaldehyde, formalin, paraformaldehyde and glutaraldehyde: what they are and what they do; *Microscopy Today* 18–12.
27. Lundblad, R.L. and Noyes, C.M. (1984) in: *Chemical Reagents for Protein Modification*; CRC Press, Boca Raton, 123–139.
28. Payne, J.W. (1973) Polymerization of proteins with glutaraldehyde; *Biochem. J.* 135, 867–873.
29. Pillai, S. and Bachhawat, B.K. (1979) Affinity immobilization and “negative” crosslinking: a probe for tertiary and quaternary protein structure; *J. Mol. Biol.* 131, 877–881.
30. Lusty, C.J. (1999) A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography; *J. Appl. Crystallogr.* 32, 106–112.
31. Janknedht, R., Martynoff, G., Lou, J., Hipskind, R., Nordheim, A. and Stunnenberg, H.G. (1991) Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus; *Proc. Natl. Acad. Sci. USA* 88, 8972–8976.

32. Pillai, S. and Bachhawat, B.K. (1977) Protein–protein conjugation on a lectin matrix; *Biochem. Biophys. Res. Commun.* 75, 240–245.
33. Banner, D.W., Kokkinidis, M. and Tsernoglou, D. (1987) Structure of the ColE1 rop protein at 1.7Å resolution; *J. Mol. Biol.* 196, 657–675.
34. Fadouloglou, V.E., Deli, A., Glykos, N.M., Psylinakis, E., Bouriotis, V. and Kokkinidis, M. (2007) Crystal structure of the BcZBP, a zinc-binding protein from *Bacillus cereus*: functional insights from structural data; *FEBS J.* 274, 3044–3054.
35. Fadouloglou, V.E., Kotsifaki, D., Gazi, A.D., Fellas, G., Meramveliotaki, C., Deli, A., Psylinakis, E., Bouriotis, V. and Kokkinidis, M. (2006) Purification, crystallization and preliminary characterization of a putative LmbE-like deacetylase from *Bacillus cereus*; *Acta Crystallogr. F* 62, 261–264.
36. Kunz, W., editor. (2010) *Specific Ion Effects*. World Scientific Publishing Co; Singapore.
37. Zangi, R. (2010) Can salting-in/salting-out ions be classified as chaotropes/kosmotropes?; *J. Phys. Chem. B* 114, 643–650.
38. Erickson, H.P. (2009) Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy; *Biol. Proced. Online* 11, 32–51.
39. Israelachvili, J. (1991) *Intermolecular & Surface Forces*; Academic Press, London.
40. Gronau, G., Qin, Z. and Markus, J. (2013) Buehler, Effect of sodium chloride on the structure and stability of spider silk's N-terminal protein domain; *Biomater. Sci.* 1, 276–284.
41. Lamba, J., Paul, S., Hasija V., Aggarwal, R. and Chaudhuri, T.K. (2009) Monitoring Protein Folding and Unfolding Pathways through Surface Hydrophobicity Changes Using Fluorescence and Circular Dichroism Spectroscopy; *Biochemistry (Moscow)*, 74, 393–398.
42. Berrya, R., Jowitta, T.A., Ferranda, J., Roessleb, M., Grossmann, J. G., Canty-Laird, E.G., Kammerer, R. A., Kadler, K. E. and Baldock, C. (2009) Role of dimerization and substrate exclusion in the regulation of bone morphogenetic protein-1 and mammalian tollid; *Proc. Natl. Acad. Sci. USA* 106, 8561–8566.
43. Li, W., Zhou, R. and Mu, Y. (2012) Salting Effects on Protein Components in Aqueous NaCl and Urea Solutions: Toward Understanding of Urea-Induced Protein Denaturation; *J. Phys. Chem. B* 116, 1446–1451.
44. Gronau, G., Qin, Z. and Markus, J. (2013) Buehler Effect of sodium chloride on the structure and stability of spider silk's N-terminal protein domain; *Biomater. Sci.* 1, 276–284.
45. Yi, F., Doudevski, I. and Regan, L. (2010) HOP is a monomer: Investigation of the oligomeric state of the co-chaperone HOP; *Protein Sci.* 19, 19–25.
46. Sampath, V., Liu, B., Tafrov, S., Srinivasan, M., Rieger, R., Chen, E.I. and Sternglanz, R. (2013) acetyltransferases from *S. cerevisiae* Hpa3-two small closely related

- Biochemical characterization of Hpa2 and Enzymology; *J. Biol. Chem.* 288, 21506–21513.
47. Wittig, I., Braun, H-P. and Schägger, H. (2006) Blue native PAGE; *Nat. Protoc.* 1, 418–428.
 48. Sreerama, N., Venyaminov, Y.S. (2000) and Woody, R.W. Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Inclusion of Denatured Proteins with Native Proteins in the Analysis; *Anal. Biochem.* 287, 243–251.
 49. Greenfield, N.J. (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions; *Nat. Protoc.* 1, 2527–2535.
 50. Compton, L.A. and Johnson, W.C. Jr (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication; *Anal. Biochem.* 155, 155–167.
 51. Munishkina, L.A. and Fink, A.L. (2007) Fluorescence as a method to reveal structures and membrane-interactions of amyloidogenic proteins; *Biochim. Biophys. Acta* 1768, 1862–1885.
 52. Ragone, R., Colonna, G., Balestrieri, C., Servillo, L. and Irace, G. (1984) Determination of tyrosine exposure in proteins by 2nd-derivative spectroscopy; *Biochemistry* 23, 1871–1875.
 53. Royer, C.A. (2006) Probing protein folding and conformational transitions with fluorescence; *Chem. Rev.* 106, 1769–1784.
 54. Dusa, A., Kaylor, J., Edridge, S., Bodner, N., Hong, D.P. and Fink, A.L. (2006) Characterization of oligomers during alpha-synuclein aggregation using intrinsic tryptophan fluorescence; *Biochemistry* 45, 2752–2760.
 55. Uversky, V.N., Li, J. and Fink, A.L. (2001) Evidence for a partially folded intermediate in alpha-synuclein fibril formation; *J. Biol. Chem.* 276, 10737–10744.
 56. Liang, Y., Du, F., Sanglier, S., Zhou, B-R., Xia, Y., Dorsselaer, A.V., Maechling, C., Kilhoffer, M-C. and Haiech, J. (2003) Unfolding of rabbit muscle creatine kinase induced by acid. A study using electrospray ionization mass spectrometry, isothermal titration calorimetry, and fluorescence spectroscopy; *J. Biol. Chem.* 278, 30098–30105.
 57. Fan, Y.X., Zhou, J.M., Kihara, H. and Tsou, C.L. (1998) Unfolding and refolding of dimeric creatine kinase equilibrium and kinetic studies; *Protein Science* 7, 2631–2641.
 58. Nakamura, S. and Kidokoro, S. (2005) Direct observation of the enthalpy change accompanying the native to molten-globule transition of cytochrome c by using isothermal acid-titration calorimetry; *Biophys. Chem.* 113, 161–168.
 59. Brogan, A.P., Widger, W.R., Bensadek, D., Riba-Garcia, I., Gaskell, S.J. and Kohn, H. (2005) Development of a technique to determine bicyclomycinrho binding and stoichiometry by isothermal titration calorimetry and mass spectrometry; *J. Am. Chem. Soc.* 127, 2741–2751.
 60. Baker, B.M. and Murphy, K.P. (1996) Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry; *Biophys. J.* 71, 2049–2055.

61. Connelly, P.R., Varadarajan, R., Sturtevant, J.M. and Richards, F.M. (1990) Thermodynamics of protein-peptide interactions in the ribonuclease S system studied by titration calorimetry; *Biochemistry* 29, 6108–6114.
62. Thao S. and Escalante-Semerena, J.C. (2011) Biochemical and thermodynamic analyses of salmonella enterica Pat, a multidomain, multimeric N(ϵ)-lysine acetyltransferase involved in carbon and energy metabolism; *MBio.* 2, 1–11, doi: 10.1128/mBio.00216-11.
63. Machius, M., Brautigam, C.A., Tomchick, D. R., Ward, P., Otwinowski, Z., Blevins, J.S., Deka, R. K. and Norgard. M.V. (2007) Structural and biochemical basis for polyamine binding to the Tp0655 lipoprotein of *Treponema pallidum*: Putative role for Tp0655 (TpPotD) as a polyamine receptor; *J. Mol. Biol.* 373, 681–694.
64. Norris, A.L., Ozen, C. and Serpersu, E.H. (2010) Thermodynamics and kinetics of association of antibiotics with the aminoglycoside acetyltransferase (3)-IIIb, a resistance-causing enzyme; *Biochemistry* 49, 4027–4035.



UNCONVENTIONAL
TARGETS OF MULTI-DRUG
RESISTANT ACINETOBACTER
JYOTI S. TOMAR
IITR

4.1. Introduction

Acinetobacter baumannii, a multidrug resistant pathogen species prevalent in the hospital acquired infections, is a serious concern for clinicians. The ability of organism to survive in the dry and abiotic surfaces for long period might be the reason for its transmission among health care setting [1–4]. *A. baumannii* has with time acquired different potent mechanisms for resistance towards majority of the commercial available drugs making it difficult to treat any *Acinetobacter* infection. Diversity of resistant mechanisms involved in the drug resistivity is already discussed in *Chapter 1*. Due to its multiple-drug resistance and with unpredictable susceptibility patterns, it becomes difficult to make the therapeutic decisions. Concerned with these facts, alternate drug targets were looked upon for preventing the *A. baumannii* menace.

As discussed in *Chapter 3*, the idea of the present study is to design new class of inhibitors for *A. baumannii* using the unconventional targets and for this reason Hpa2 (Histone acetyltransferase) and TAG (a DNA repair enzyme) enzymes were characterized. In this chapter, recombinant production of the enzyme TAG and its biophysical characterization are discussed.

Integrity and stability of the genome is crucial to prevent any variation in the conformity of information embedded as nucleotide sequences [5–7]. However, the genome integrity gets constantly damaged through different mechanisms, which leaves the DNA prone to several dangerous mutations. Expression of bacterial genes is also a highly regulated process with primary gene sequence as the principal regulator of its expression. This statement indicates the significance of the genome sequence stability. Different mechanisms are studied to be causing damage to the DNA *viz.* highly reactive metabolic by products which react with DNA causing its damage [8, 9], highly reactive DNA bases which decay over time, mostly through the hydrolysis resulting in the abasic site [10], deletion or mis-incorporation of DNA bases during replication and also damage through exogenous agents *e.g.*, occurrence of mutations due to deletion or mis-incorporation of DNA bases during industrial chemical vinyl chloride [11–13]. To solve this problem, the organism evolved several DNA repair mechanisms.

The DNA repair primarily addresses mismatches created during damage [13–16]. These damages might occur in the form of single base mismatches, single base loops, or insertion and deletion of the loops [17, 18]. The well known pathway to repair these

mismatches are nucleotide excision repair (NER) and base excision repair (BER). Generally, NER process is initiated after the recognition of a disrupted base-pair causing distortion of the DNA helix and the distortions which do not affect the helix are addressed by BER. However, base excision repair pathway corrects the non-helix-distorting mismatches and this pathway is initiated by lesion-specific enzymes known as DNA glycosylases. Generally substrate specificities of the enzymes glycosylases are moderately overlapping. So, more than one enzyme play role in the correcting the lesions. The function of the glycosylase enzyme and steps involved in the BER pathway are described in *Chapter 1*. In prokaryotic cells, two DNA glycosylases, 3-methyladenine DNA glycosylase II (AlkA) and 3-MeA DNA glycosylase I (TAG) are found to cleave the glycosidic bond of alkylated purine bases in DNA *via* hydrolysis. 3-Methyladenine DNA glycosylase II (AlkA), has been studied in free as well as in bound form and structural analysis reveals it to be a member of HhH family. It cleaves the modified/mispaired adenine and guanine using water as activated nucleophile. The conservation of structural fold suggests that this scaffold has evolved for binding to nucleic acids during the early stages of evolution.

The second enzyme, 3-MeA DNA glycosylase I (TAG) is expressed constitutively in the bacterial cells. Due to the absence of sequence similarity with other DNA glycosylases, its phylogeny is not well understood. It does not possess sequence as well as structural similarity with its structural and functional cousin AlkA. Owing to its significance in bacterial functioning and its absence in humans it is an excellent chemotherapeutic target to be studied.

4.2. Techniques used for biophysical characterization of TAG

In addition to biophysical techniques described in *Chapter 3*, here we discuss about the techniques which are used to study the enzyme TAG.

4.2.1. HPLC-based assay of glycosylase enzyme

HPLC has been for long used to monitor the enzymatic activities. It is widely used to study the kinetics of different enzymes. HPLC-based assays are given an edge over other existing methods due to: (i) its utmost accuracy, specificity, sensitivity, as well as reproducibility, (ii) user-friendly software enabling the easy control over sampling and processing of data and (iii) host of detectors provide a means to virtually detect any compound. Due to these factors,

HPLC-based assay for glycosylase enzyme TAG was developed which is being discussed here in detail.

Enzyme activity is checked by incubating the control DNA (cellular methylated DNA) (50 μM) with enzyme TAG (2 μM) and amount of free 3mA released as a product was quantified from peak area. The control DNA was added to enzyme TAG with a final volume of 500 μL and the reaction mixture was incubated at 37 °C till maximum substrate was consumed (120 min). Prior to HPLC analysis, the reaction mixture was subjected for the neutral thermal hydrolysis at 90 °C. The product of enzyme activity (3mA base) was separated by reversed-phase HPLC. The control DNA, TAG enzyme and 3mA were also analyzed separately under the same experimental conditions to get the retention time.

4.3. Materials and Methods

4.3.1. Chemicals

Chemicals used in these studies are of analytical grade and most of the chemicals were purchased from Fluka (Buchs, Switzerland), Sigma-Aldrich (Deisenhofen, Germany), Merck (Dramstadt, Germany), and Himedia. PCR purification, gel extraction and plasmid miniprep Kits were purchased from Qiagen. Agarose gel electrophoresis unit and gel imager unit were procured from Biorad and chemicals used in the DNA gel electrophoresis such as ethidium bromide, agarose, and gel loading dye were purchased from Biochem. Isopropyl β -D-1-thiogalactopyranoside used to induce protein expression was purchased from Biochem. The reagents used in the sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE) and in the Bradford assay were purchased from the Bio-rad. Columns for gel filtration; HisTrap HP and HiLoad 16/600 Superdex 200 were purchased from GE healthcare. The amicon centrprep concentrators with 3 kDa cut-off and dialysis tubing spectra/por MW 3,500 were purchased from the Millipore.

4.3.2. Bacterial strains and vectors

Escherichia coli DH5 α and BL21 (DE3) strains and cloning and expression plasmids were purchased from the Novagen (USA).

4.3.3. Molecular biology enzymes

Restriction endonuclease enzymes (*Nde*I, *Xho*I), bovine serum albumin, reaction buffers, T4

DNA ligase, DnaseI, Rnase A, Pfu Turbo polymerase and Taq polymerase were purchased from New England Biolabs (NEB). Bovine serum albumin (BSA) and reaction buffers were from NEB. Enzymes utilized during purification, lysozymes and deoxyribonuclease I (DNase I) were purchased from the Sigma–Aldrich.

4.3.4. Oligonucleotides for PCR amplification

The following set of DNA oligonucleotides were used for the PCR amplification of TAG gene. The *Nde*I and *Xho*I restriction sites were introduced in forward and reverse primers, respectively. The oligonucleotides were supplied by Biosciences, Bangalore, India. The sequences of primers are presented in 5' to 3' orientation.

Forward Primer:

5' GAT TCT CAT ATG ATGAAACTCAACGCTGCGGATG 3'

Reverse Primer:

5' GAT TCT CTC GAG ATGAAGCTTTAAATTGACAATCATTTC 3'

4.3.5. Culture media

LB (Luria–Bertani) broth and LB agar for growing the *E. coli* cells were procured from Merck speciality chemicals as dried granules. The LB agar consists of 1% [w/v] tryptone, 0.5% [w/v] yeast extract, 1% [w/v] NaCl. The LB culture media consisting of tryptone 1% [w/v], yeast extract 0.5%, NaCl 1% [w/v] was sterilized by autoclaving at 121 °C for 15 min at 15 psi pressure. Antibiotics kanamycin and chloramphenicol were purchased from Himedia chemicals, India, were added into the culture for selective growth. Antibiotics; kanamycin (Stock solution: 100 mg/mL in *dd*H₂O, working concentration: 50 µg/mL), chloramphenicol (stock solution: 100 mg/mL, in absolute ethanol, working concentration: 35 µg/mL).

4.3.6. Gene cloning and vector construction

The oligonucleotides for PCR amplification were purchased from Integrated DNA technologies Inc. (Coralville, IA). DNA gel extraction and PCR purification kits were ordered from Qiagen Inc. XL1-Blue strain of *Escherichia coli* was used for transformation and amplification of plasmids. Restriction enzymes *Nde*I and *Xho*I were purchased from New England BioLabs, Inc. *E. coli* BL21 (DE3) expression strain was used as a host for protein overexpression and pET-28a vector were obtained from Novagen Inc. Madison, WI, USA.

The genomic DNA of *A. baumannii* was used as the template for gene amplification using the two primers (forward) 5' GATTCTCATATGATGAAAACCTCAACGCTGCGGATG 3' and (reverse), 5' GATTCTCTCGAGTTATGAAGCTTTAAATGACAATCATTTTC 3'. After amplification, PCR purification kit was used to purify the PCR product which was then subjected to double digestion with *NdeI* and *XhoI* enzymes. The pET-28a plasmid was also treated with the same set of restriction enzymes at 37 °C for 1 h. The digested fragments separated on 1% low melting agarose gel were excised after completion of the run. Subsequently, fragments were eluted from the excised band using DNA gel extraction kit and ligated in the presence of T4 DNA ligase enzyme. The ligated PCR-plasmid product was transformed into the *E. coli* XL1-Blue competent cells by the heat shock method [19]. Colonies obtained by selective growth of the transformed cells on Luria–Bertani agar plates containing 50 µg/mL of kanamycin were picked up and grown overnight at 37 °C. Cloned plasmids were isolated using a MiniPrep plasmid isolation kit and screened for the presence of the *TAG* gene by restriction enzyme digestion. The integrity of the resulting plasmid, pET-28a-6xHis-TAG, was confirmed by sequencing in both directions using T7 promoter and T7 terminator primers. Further, expression and solubility optimization experiments were done using the pET-28a-6xHis-TAG plasmid.

4.3.7. Expression optimization of the pET-28a-6xHis-TAG

Plasmid pET-28a-6xHis-TAG having the *TAG* gene of *A. baumannii* was transformed into *E. coli* expression strain BL21 (DE3). The expression and solubility of His-tagged TAG protein were tested by growing 5 mL bacterial cultures in LB broth having appropriate antibiotics concentration. The overnight cultures were then diluted to 1/100 with fresh LB broth supplemented with suitable concentration of antibiotics and incubated at 37 °C till A_{600} becomes 0.6. The cultures were then induced with different concentrations of IPTG (1.0, 0.5, 0.25, 0.1 mM) and grown overnight at different temperatures (37, 25, and 18 °C) with shaking at 225 rpm. Cells were harvested by centrifugation at 7000 rpm for 6 min at 4 °C. The cell pellets were resuspended in 5 mL of 50 mM Tris buffer, pH 7.6, and the cells were disrupted using a French press. Then expression and solubility confirmation were done by analyzing this induced lysed cell supernatant and the pellet on 15% sodium dodecylsulfate poly-acrylamide gel electrophoresis (SDS–PAGE) with uninduced control lysed cell supernatant and pellet.

4.3.8. Purification of pET-28a-6xHis-TAG

Purification of His-tagged TAG protein was performed at low temperatures (~ 4 °C). The cell pellets obtained from 1-liter culture were resuspended in 25 mL of buffer A (50 mM Tris-HCl pH 7.6, 250 mM NaCl, 2 mM EDTA and 5% glycerol) and disrupted using a French press. To separate the protein present in supernatant from the cell debris, lysate was centrifuged at 15000 rpm for 45 min at 4 °C. The supernatant was incubated with Ni-NTA column, pre-equilibrated with buffer A containing 15 mM imidazole. Elution was performed by running a gradient of 500 mM of imidazole and the fractions were analyzed by SDS-PAGE. Fractions containing pure TAG were pooled and dialyzed for 3–4 h against 1 L of dialysis buffer [20 mM Tris-HCl (pH 7.6), 100 mM NaCl, 5% glycerol] containing 2 mM EDTA. Purity and oligomeric state of protein is confirmed by the size exclusion chromatography. Purified protein was concentrated using an Amicon Ultra-3 kDa concentrator and stored at –80 °C.

4.3.9. Protein concentration determination

The protein concentration determination was performed using Bradford reagent provided by Bio-Rad [20]. The protocol suggested by the supplier was followed for the assay. The 5 µL of concentrated protein sample was mixed with 795 µL of purification buffer. The 200 µL of Bradford reagent was added to make final volume 1 mL and incubated for 1 min. The protein concentration was determined using Perkin Elmer Lambda 25 UV–visible spectrophotometer by interpolating the absorbance reading at 595 nm into a BSA standard curve.

4.3.10. TAG molecular weight determination

The 5 enzymes comprising the standard curve in gel filtration analysis, (ribonuclease A, carbonic anhydrase, ovalbumin, thyroglobulin, aldolase and aprotinin), blue dextran were purchased from GE healthcare. The gel filtration column 16/600 Superdex 200, was also purchased from GE Healthcare. Other chemicals were purchased from Merck India or Sigma-Aldrich. In order to determine the oligomeric state of the TAG in the solution, an analytical 16/600 Superdex 200, column was run with running buffer (50 mM Tris pH 7.5, 100 mM NaCl, 5% glycerol, 0.5 mM DTT) at 4 °C with a flow rate of 0.5 mL/min. The void volume was first determined by injecting 50 µL of blue dextran (conc. 8 mg/mL) and establishing its elution volume (40 mL). A series of standards (1 mg each) were dissolved in 1000 µL of running buffer, of which 500 µL was loaded on the column. The retention volume of each

standard was recorded and their K_{av} values were calculated using the equation, $K_{av} = (V_e - V_o)/(V_t - V_o)$, where V_e is the elution volume, V_o is the void volume, and V_t denotes the total volume. A plot of K_{av} as a function of $\log(MW)$ produces a standard curve. Purified TAG was diluted to 3 mg/mL in buffer A, and 500 μ L of which was injected on the column and run under the conditions as described above. Retention times of the peaks were recorded, and the MWs of different standards (ribonuclease A, carbonic anhydrase, ovalbumin, thyroglobulin, aldolase and aprotinin) were determined using the above equation and by fitting in the standard curve (Figure 1). This curve enables the estimation of molecular size value of TAG protein. The de-termination of molecular weight of protein TAG was estimated in the physiological condition.

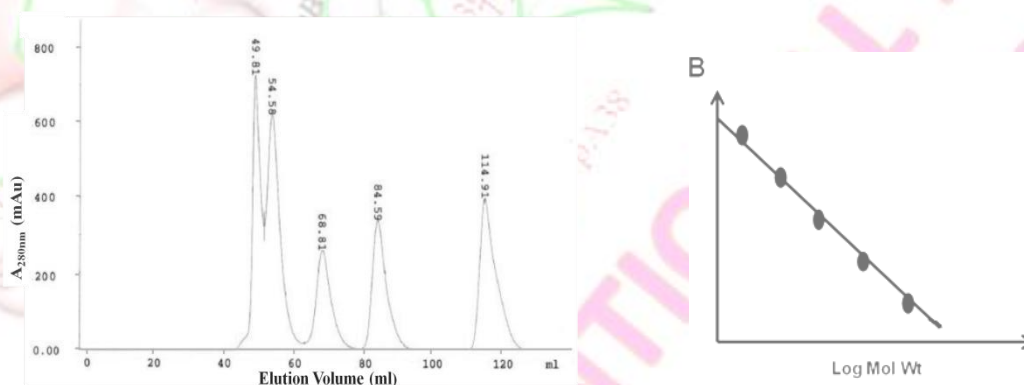


Figure 1. Determination of molecular mass using gel filtration chromatography.

4.3.11. Circular dichroism

The circular dichroism (CD) spectra were measured on Chirascan Circular Dichroism Spectrometer (Applied photophysics Ltd, Surrey KT22 7PB, UK). The scan speed was 50 nm min^{-1} with 1 nm bandwidth, 0.5 nm wavelength steps and an average time of 3.0 sec. For the far-UV CD measurement, wavelength range of 190 to 260 nm and for the near-UV CD measurement, the wavelength range of 260 to 320 nm were taken with three accumulations in a quartz cell with nominal path length of 1 mm. Protein sample is dialyzed for 4 h in CD buffer containing 25 mM sodium phosphate buffer of pH 7.6 and 100 mM NaCl. After dialysis, the contents were centrifuged at 14000 rpm for 30 min and then filtered through 0.1 to 0.2 micron filters to reduce light scattering. Protein samples at concentration of 1 mg/mL were used for recording spectra and the same buffer was used as the blank. For each sample, collected scans were averaged, and then the baseline corresponding to the buffer was

subtracted to obtain the final spectra. The final spectra were analyzed using the program CDNN [21]. Effects of variable concentrations of substrate (3mA) and Zn binding on the secondary and tertiary structure of TAG were determined by incubating the protein with 10, 50, 100, 200, 300, 400, 500 and 600 μM 3mA and 50, 100 μM of ZnCl_2 for 4 h at 4 $^\circ\text{C}$ followed by the far-UV CD and near-UV CD measurements.

4.3.12. Isothermal titration calorimetry binding assay

As discussed in *Chapter 3*, a syringe containing ligand is titrated into a cell containing protein sample. When the protein interacts with the ligand, heat released or absorbed is directly proportional to the amount of binding that occurs. To determine the binding parameters of TAG with substrate 3-methyladenine and zinc chloride, isothermal titration calorimetry (ITC) assay was performed using an ITC-200 Microcal calorimeter (GE Healthcare). All the experiments were performed at 20 $^\circ\text{C}$ temperature. The pure TAG protein and ligands were prepared in 25 mM Tris buffer pH 7.6. The cell was filled with 225 μL of 50 μM of protein and the syringe was loaded with 60 μL of 3 mM ligands (3-methyladenine and Zn^{2+}) as binding partner, 20 injections of 2 μL ligand, each was injected into the cell at a stirring speed of 2000 rpm. The time interval between two injections was set to 120 sec to ensure optimum mixing and measurement of change in reaction enthalpy. The titration data of ligand with protein-free buffer was used as reference. The data analysis was done using Origin software version 7.0.

4.3.13. Spectrofluorimetric studies

Fluorescence spectra were recorded on Fluorolog-3 Spectrofluorimeter LS55 (make HORIBA Jobin Yvon Spex®). Extrinsic fluorescence emission spectrum in the range of 350–500 nm was recorded upon excitation at 350 nm. The slit widths were typically 5 nm and sample volume varied 1.5 mL (10 mm cuvette) depending upon the concentration of protein required to obtain the satisfactory signal to noise ratio. Before the measurement, sample cells were equilibrated for 90 sec at each temperature. Binding experiments were done at a TAG concentration of 2 μM in the presence of binding buffer containing 20 mM Tris pH 7.6 and 100 mM NaCl. The binding affinity of 3mA with TAG was determined by following the Trp fluorescence quenching. The fixed concentration of TAG (2 μM) was titrated with increasing amounts of 3mA (1–500 μM) using a 2.5 mL quartz cuvette with a magnetic stir bar ($T = 20$

°C). The emission spectra were recorded in the range of 290–400 nm. In order to determine the K_d , a titration experiment was performed by monitoring the fluorescence change at 350 nm. To have an estimate of the K_d , the binding data was fitted using the following equation and the software program Sigma Plot

$$F = f_1 \cdot P_o + \frac{1}{2} (f_2 - f_1) \left[(P_o + L_o + K_{Diss}) - \sqrt{(P_o + L_o + K_{Diss})^2 - 4 \cdot P_o \cdot L_o} \right]$$

where f_1 denotes fluorescence signal of the protein with zero concentration of substrate *i.e.*, 3mA, f_2 is the fluorescence signal of the protein when it is fully saturated with the substrate. P_o ; concentration of protein TAG and L_o ; concentration of the ligand and K_{Diss} is the binding constant. The thermal denaturation was performed to observe the effect of temperature (20–75 °C) on the TAG.

4.3.14. DNA glycosylase assay using HPLC

To measure glycosylase activity, control DNA was incubated with enzyme 2 μ M TAG (present in 20 mM Tris, pH 7.5 and 50 mM NaCl) at 37 °C. The final substrate concentrations was 50 μ M (1 : 25 ratio). The reaction mixture was kept for 20–120 min. After incubation, the reaction mixtures were subjected to neutral thermal hydrolysis at 90 °C for 10 min to release free 3mA. 10 μ L of cold 1 M HCl was used to precipitate depurinated DNA. Then 20 μ L of reaction mixture was injected in the HPLC column. The fraction of 3mA was calculated from the area under the peak. Before the product was analyzed, HPLC analysis of the pure 3mA was performed using the same elution solvents and experimental parameters. The Bases released into the supernatant were separated by reversed-phase chromatography (Beckman Ultrasphere C₁₈ column, 25 cm \times 4.6 mm, 5 μ M particle size) using acetonitrile/water as mobile phase at a flow rate of 0.5 mL/min. The 3mA adduct was detected by a UV-detector.

4.4. Results and discussion

4.4.1. Cloning of TAG gene

To get the recombinant soluble TAG protein, gene encoding 3-methyladenine glycosylase-I was cloned into pET-28a-TEV, an *E. coli* based expression vector with an engineered N-terminal 6xHis-tag and TEV protease cleavage site. The gene synthesis of the TAG protein was done from the Biolinkk. The forward and reverse primers with the restriction

endonucleases sites of *NdeI* and *XhoI* were used for the PCR amplification. The vector DNA and PCR product were double digested with *NdeI* and *XhoI* restriction enzymes and T4 DNA ligase was used to ligate digested PCR product into digested pET-28c-TEV vector. The ligated product was then transformed into *E. coli* DH5 α cells. The plasmid DNA isolated from transformed cells was subjected to double digestion with *NdeI* and *XhoI*, digestion with these R.E. furnished the fragment of desired gene length (approx. 450 bp) (Figure 2). The automated DNA sequencing further confirmed the specificity of cloned DNA fragment.

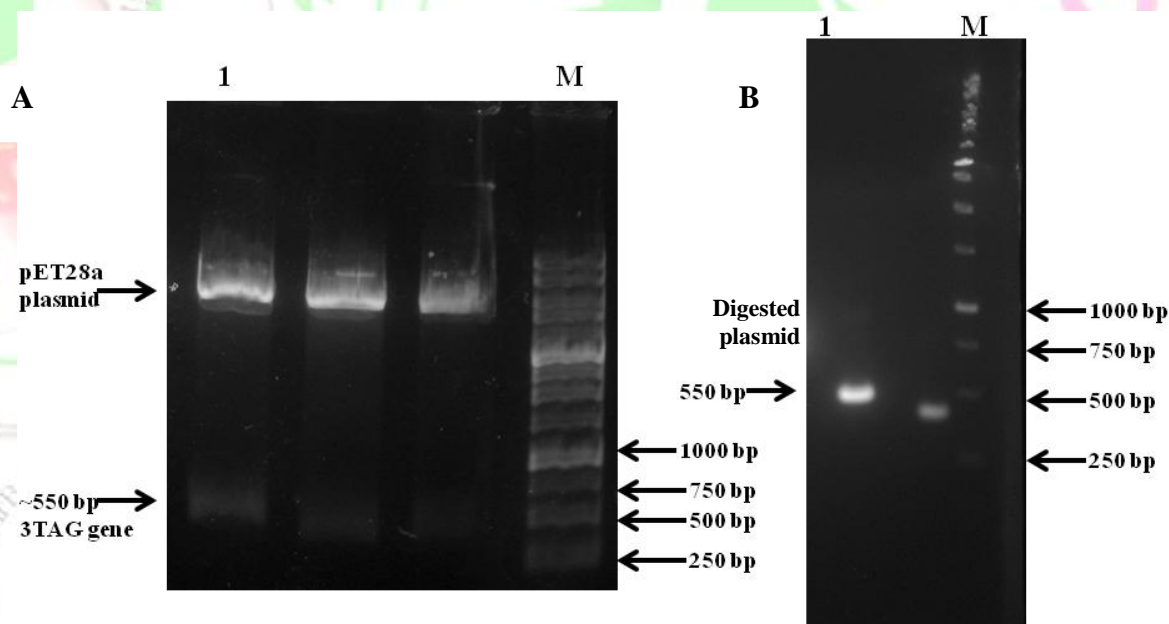


Figure 2. Amplification and cloning of TAG gene. Agarose gel electrophoresis (A) Recombinant plasmids digested with *NdeI* and *XhoI*: Lane 1: Restriction digestion of screened colony, Lane M: 1kb DNA ladder *geneDirex*. (B) PCR product: Lane1: PCR amplification of 3MAG gene from genomic DNA of *A. baumannii*, Lane M: 1kb DNA ladder *geneDirex*.

4.4.2. Optimization of condition for expression and purification of TAG protein

The expression of His6-tagged-TAG protein at 37 °C in *E. coli* BL21 (DE3) led to the formation of inclusion bodies regardless of the IPTG concentration used for induction. The expression was optimized by varying temperature (37, 25 and 19 °C), the IPTG concentration for the induction of expression (0.05, 0.1, 0.2, 0.4, 0.5, 0.7 and 1.0 mM), and pH of the media (6.0, 6.5, 7.0, 7.5, 8.0). Analysis of SDS-PAGE electrophoresis showed that expression of recombinant *A. baumannii* TAG protein in soluble form with the expected molecular mass

(~21 kDa) is obtained by growing the transformed *E. coli* cell at 18 °C for ~ 16 h after induction with 0.7 mM IPTG in the LB medium having pH 6.5 (Figure 3A).

The protein was purified by Ni-affinity chromatography where an imidazole gradient was used for eluting bound proteins. His-tagged bound protein was eluted from the column at an imidazole concentration of ~200 mM. The purified TAG protein fractions were collected and to evaluate purity, the collected fractions were analyzed on a 12% SDS-PAGE gel which was stained with Coomassie blue. It is estimated that the purity of the protein fractions are >95%, corresponding to ~21 kDa band (Figure 3B). Fractions having TAG protein were subjected to a three-step dialysis in phosphate buffer for the buffer exchange. Purified protein was the oligomeric state of the protein in the solution, gel filtration chromatography is performed using 16/600 Superdex 200 prep-grade gel filtration column. From the column, TAG is eluted as a single peak at volume corresponds with molecular mass of ~21 kDa, which confirms the presence of TAG protein as monomer in aqueous solution (Figure 4). Fractions having pure protein are concentrated to 1 mg/mL for the biophysical experiments and are stored at -80 °C.

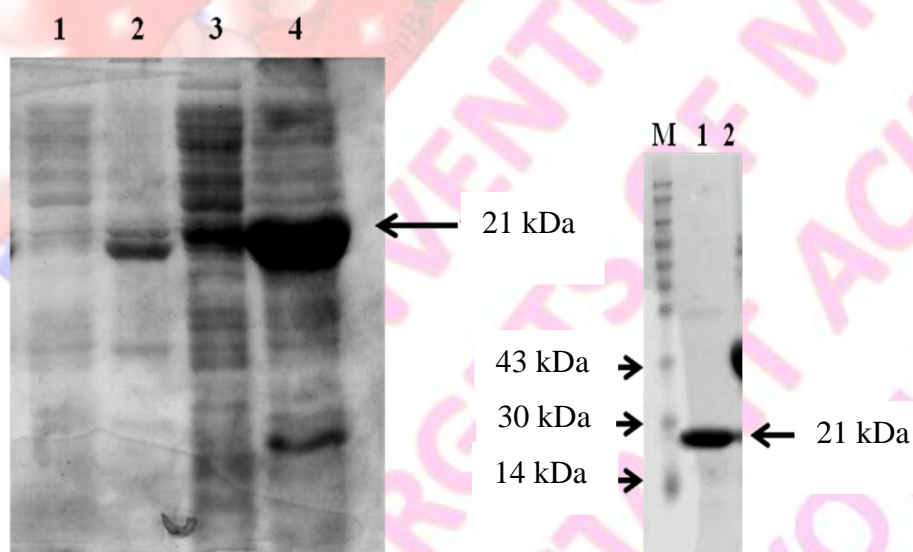


Figure 3 (A) Expression of TAG observed on 15% SDS-PAGE: Lane1: Uninduced soluble fraction, Lane 2: Uninduced insoluble fraction, Lane 3: Induced soluble fraction, and Lane 4: Induced insoluble fraction. (B) Purification of TAG enzyme observed at 15% SDS-PAGE. Lane M: Pre-stained protein ladder *geneDirex*, Lane1: Purified TAG.

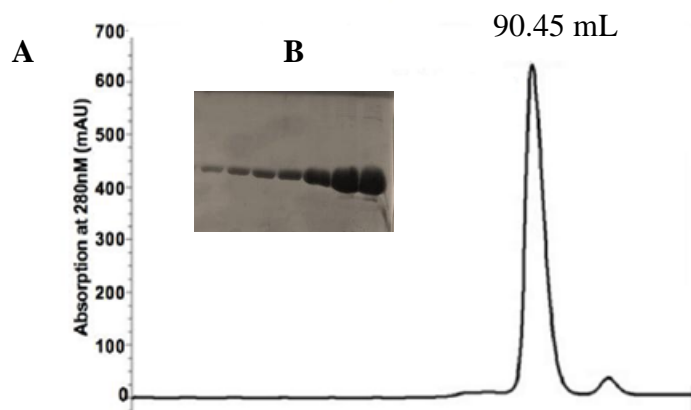


Figure 4. Purification and molecular mass determination of TAG (A) Affinity purification of TAG showing purified protein in 15% SDS-PAGE. (B) Size exclusion chromatography profile of TAG analysis of major peak (using HiLoad 16/600 Superdex 200 column, $V_0 = 45$ mL and flow rate 0.5 mL/min).

4.4.3. Secondary structure and substrates binding analysis using CD spectroscopy

The CD is acknowledged as a valuable biophysical technique for examining the protein structure in solution. It is considered as an excellent method for the speedy evaluation of secondary structure, conformational changes and binding properties of proteins. When amide chromophores, the building blocks of protein backbone, are aligned in arrays, their optical transitions split into multiple transitions due to exciton interactions. These interactions yield distinct CD spectra for different types of secondary structure and these characteristic spectra have been utilized in the quantitative estimation of secondary structure and monitoring of protein folding and unfolding. In addition to several aspects of TAG discussed in previous sections, the secondary structure and the binding properties of TAG with substrates are herein described with the help of CD experiments to support the results obtained in the current *in silico* structural studies (Section 2.5). Secondary structure of TAG was studied using far-UV circular dichroism and representative CD spectra of TAG are presented in Figure 5.

The characteristic minima (209 and 222 nm) of α -helical proteins are clearly visible in the CD spectrum (native) of the TAG. The spectral data was analyzed and deconvoluted by the program CDNN [22] which yields a secondary structural content of $62.2\% \pm 10\%$ α -helix, $3.1\% \pm 1\%$ β -sheet, $6\% \pm 3\%$ β -turn, and $28.7\% \pm 7\%$ random coil. This is very close to secondary structural content predicted theoretically.

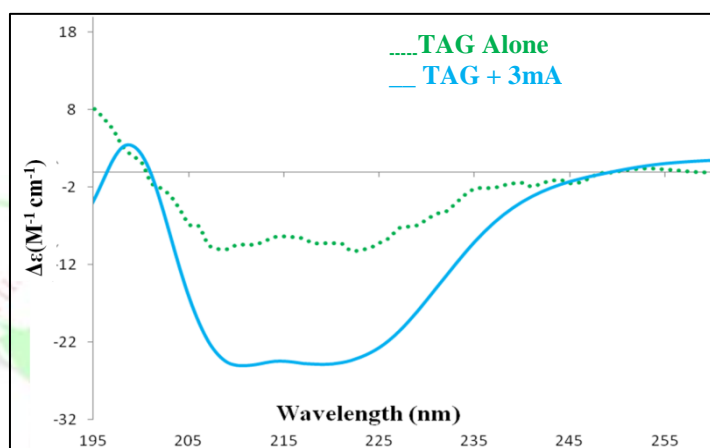


Figure 5. Far-UV circular dichroism analysis of TAG alone (green) and after substrate 3mA binding (blue) in 25 mM sodium phosphate buffer at 20 °C.

Structural changes in protein caused by the binding of ligands are essential part of the mechanism of action which also regulates its biological activity. The CD provides an experimentally convenient mean of detecting such changes at different spectral regions. CD was also used in case of TAG to access the ligand concentration over which spectral changes take place and to measure the extent of the change. On addition of the substrate 3mA to the enzyme TAG, significant change in the intensity of CD spectra, such as increase in the signal around 222 nm as well as at 208 nm was noticed (Figure 5). The band at 222 nm can be attributed to strong H-bonding environment of α -helices and is independent of their length. From experimental studies on proteins of the same family, HhH motif majorly contributes towards the active site. Due to this reason, various concentrations of substrate binding at CD signal 222 nm were monitored (Figure 6A). This shows how binding may induce protein to be more helical *i.e.*, binding increases the compactness in the three-dimensional structure of the TAG. Analysis through CDNN program shows that after substrate binding, helical content increases from 62–91% \pm 10%. The ellipticity at 222 nm as a function of increased substrate concentration showed that the ellipticity decreased \sim 6 times on saturation as compared to native TAG, and shape of graph represents the characteristic graph of protein refolding. As large increase in the helical content was observed on substrate binding which may not be only due to substrate binding, because the recombinant enzyme TAG may exist as the molten globule state in solution and binding of substrate may induce protein folding to craft proper-shaped active pocket. The change in ellipticity can be used to determine the substrate binding constants.

The changes in shape and magnitude in near-UV CD spectra of protein depend up on the nature of each type of aromatic amino acid present, and their environment (H-bond, polar groups). It is important to yield structural insight about a particular residue of the protein. The change in tertiary structure is majorly caused by the substrate binding. This is clear from near-UV CD spectra (Figure 6B) that the environment of pocket is modified due to substrate binding [23].

On titration with substrate, a significant change of positive shift in the near-UV CD spectrum was identified, especially in 270–285 nm region of the Tyr residue. From these spectral changes, it can be inferred that local tertiary structure of Tyr in the protein is affected upon substrate binding. The current molecular docking analysis indicates the interaction of substrate in the binding pocket resulted in the formation of H-bond with the residues Tyr15

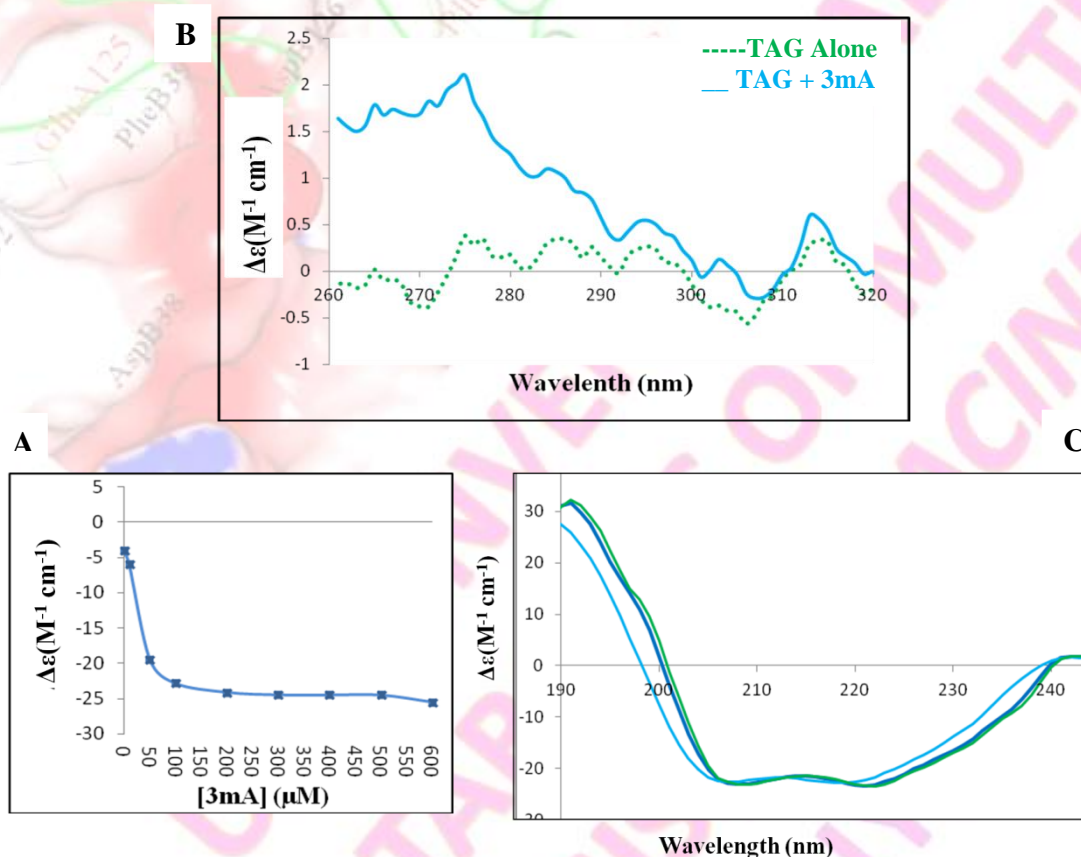


Figure 6. (A) Change in ellipticity with addition of 10, 50, 100, 200, 300, 400, 500 and 600 μM of substrate 3mA monitored at CD signal 222 nm. (B) Near-UV CD spectra of TAG alone (green) and after substrate 3mA binding (cyan) at 20 °C. (C) Effect of Zn-metals on the conformational stability of TAG, CD spectra of TAG in 25 mM sodium phosphate buffer, with addition of 50 (blue), 100 μM (cyan) ZnCl_2 at 20 °C ZnCl_2 .

and Tyr18 (Figures 5B and 5D; Section 2.5.4); similarly, the spectral changes in the tyrosine region attest that the binding of 3mA affects the local tertiary structure of Tyr aromatic side chains significantly. We presumed that Zn binding by TAG would kick off conformational changes on TAG and hence the effect of Zn(II) addition on the CD spectra of TAG was measured. Slight perturbation of the CD spectrum was observed along with increase in the ellipticity as compared to native (Figure 6C).

This change in spectral shape and ellipticity indicate that binding of zinc resulted in slight reorganization in the secondary structure affecting stability of overall 3D structure of TAG. Stivers and his co-workers stated that zinc snap motif allows the TAG family to efficiently stabilize the HhH structure without additional protein scaffolding present in closely related families [24, 25].

4.4.4. Isothermal titration calorimetry binding assay of TAG with 3mA and Zn²⁺

Titration of TAG protein with the substrate 3mA produces a typical sigmoid binding isotherm. Analysis of the binding parameters indicated that binding of 3mA to TAG is driven by favourable enthalpy change ($\Delta H = -12.6 \pm 0.4 \text{ kcal mol}^{-1}$) and displayed high affinity ($K_d = 63 \pm 0.2 \text{ }\mu\text{M}$) (Figure 7A), which is comparable with the binding affinity obtained by the Stivers and co-workers for TAG enzyme of *E. coli* [26]. Binding of 3mA to TAG is not favoured by the entropy term ($-T\Delta S = 9.1 \text{ kcal mol}^{-1}$); the ΔG value is calculated using temperature 298 and it is found to be $-3.5 \text{ kcal mol}^{-1}$ indicating that binding is exothermic process. One monomer of TAG binds one molecule of 3mA ($n = 0.934$). The binding isotherm of Zn²⁺ was characterized by strong binding affinity for the TAG. The dissociation constant determined for TAG ($K_d = 266 \pm 4.1 \text{ }\mu\text{M}$) was higher than that for 3mA.

The enthalpy change ($\Delta H = -8.4 \pm 0.9 \text{ kcal mol}^{-1}$) favours the binding of Zn²⁺, and its binding is opposed by entropy term ($-T\Delta S = -4.77 \text{ kcal mol}^{-1}$) and the calculated ΔG value (Zn²⁺; $3.7 \text{ kcal mol}^{-1}$) indicates that Zn²⁺ exhibits thermodynamically favourable binding (Figure 7B). It is known from the literature [27] that Zn²⁺ possesses one binding site near to the N-terminal of TAG. However, higher number of binding sites ($n = 3.8$) suggested the presence of the nonspecific binding. We also performed the binding of adenine with the protein but it was found that adenine does not exhibit the binding with the protein.

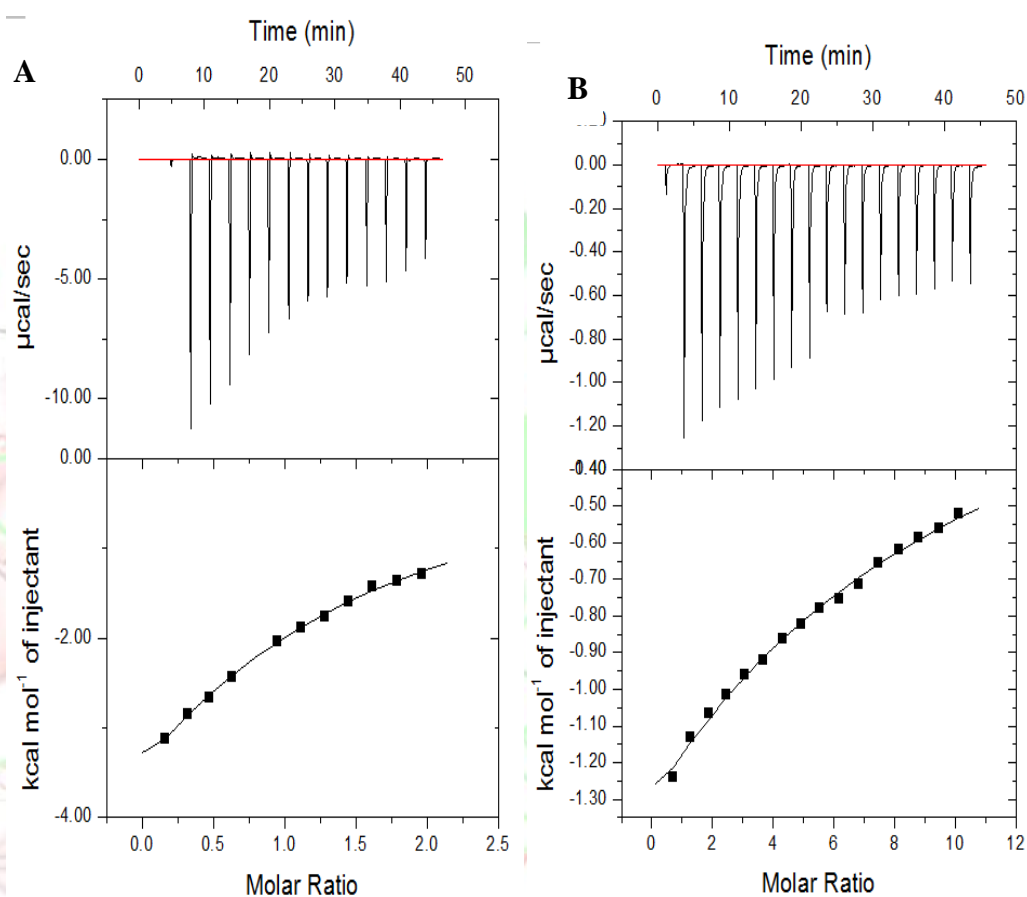


Figure 7. Binding of TAG with substrate 3mA and Zn²⁺. TAG was titrated with (A) 3mA and (B) Zn²⁺, at a temperature of 20 °C. Upper panels correspond to titration kinetics. Lower panels show the integrated binding isotherms. Molar ratio refers to protein monomer. The binding enthalpy (ΔH) and dissociation constant (K_d) were obtained by non-linear regression of the integrated data to a one-site binding model.

4.4.5. Fluorescence binding assay

The effect of temperature on the TAG was studied using the intrinsic fluorescence (Figures 8A and 8B). The emission of the aqueous tryptophan was recorded between 290–450 nm after exciting at 290 nm. Analysis of the results of thermal denaturation shows that protein is stable up to 35 °C temperature. Denaturation of protein starts after 35 °C, and at the temperature 55 °C almost 50% of the protein denatures. Denaturation curve is plotted as the intrinsic fluorescence intensity maxima F_{350} vs. temperature (Figure 8B). Analysis of denaturation curve indicates that denaturation of the TAG is a two-step process; however, unlike Hpa2 no major shift in the maxima were found. The heating of the sample above 65 °C produces the

same amount of fluorescence which means that the protein is completely denatured and so further heating does not have any effect on the intrinsic fluorescence. The thermal denaturation of TAG is also performed and extrinsic fluorescence is measured using the ANS as the probe. Analysis of obtained spectra is given in [Appendix IV-1](#). Similar pattern for TAG denaturation attests to the intrinsic denaturation results.

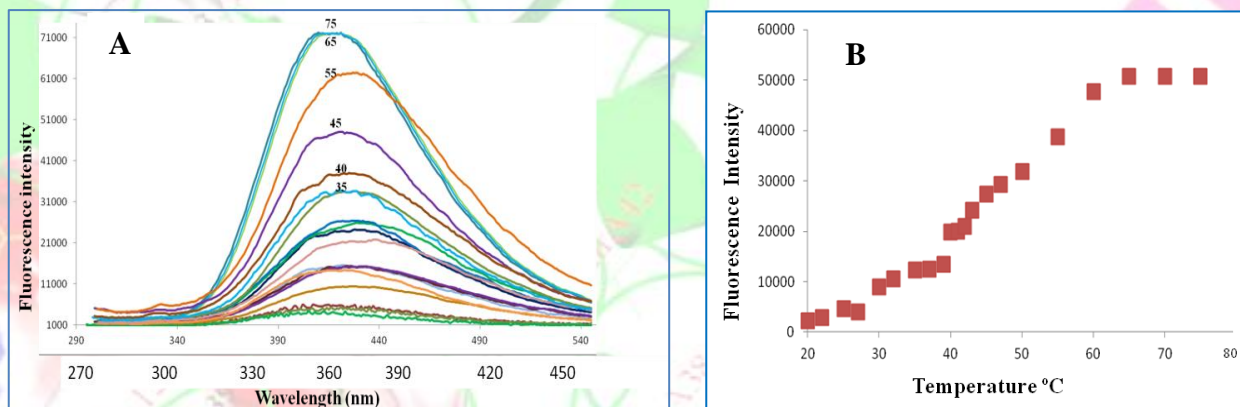


Figure 8. Thermal denaturation of the TAG (A) Intrinsic fluorescence of TAG at different temperatures and (B) Intrinsic fluorescence at wavelength 350 vs. temperature; curve was fit to the data assuming a two-state unfolding model.

The binding parameters of substrate 3mA were investigated using the neutral pH condition as described in *Section 4.3.13*. Titration of 3mA resulted in decrease of the intrinsic fluorescence intensity at 330 nm [Appendix IV-2](#). Fluorescence intensity is plotted as the function of substrate concentration and is depicted in [Figure 8C](#).

Binding constant is determined using the equation described in the *Section 4.3.13* and the dissociation constant value was found to be $63 \pm 5 \mu\text{M}$, which is in good correlation with ITC and literature data [26]. Binding of the unmodified adenine was also investigated and to no change in the spectra was observed as shown by ITC, indicating towards the fact that TAG has a significant binding affinity as well as specificity for 3mA.

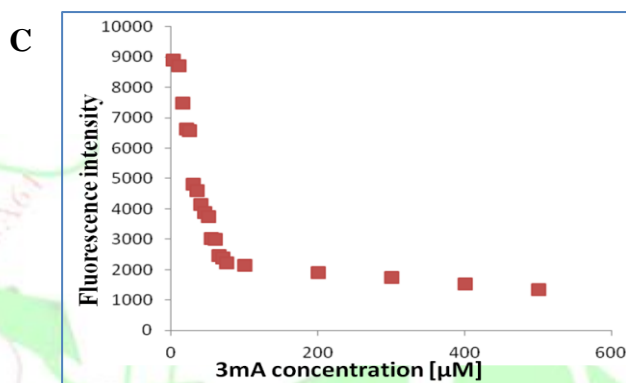


Figure 8C. Tryptophan fluorescence studies of TAG binding to 3mA, TAG shows a 65% decrease in its Trp fluorescence upon binding 3MeA (square box) but not the normal base adenine (Appendix IV-3). K_d for 3mA determined from these data is 63 μM .

4.4.6. DNA Glycosylase assay using the HPLC

For the determination of TAG activity, HPLC chromatography was performed [27–30]. The time dependent HPLC analysis of the reaction mixture was done to examine the enzyme activity. Elution is carried out with suitable solutions (water 15% : acetonitrile 80%) and product formed is qualitatively analyzed. Firstly, the control DNA and TAG protein were subjected for the HPLC analysis and a peak at retention time of 2.93 min corresponding to control DNA was obtained (Figure 9A). However, HPLC analysis of TAG produces three elution peaks, at 4.13 (major) 3.2 (minor) and, 8.18 (minor) min (Figure 9B). Minor peaks might be due to small impurity present in the protein sample.

The reaction mixture was incubated for 15 min and after incubation, 20 μL of it was injected to the column. At first, the sample runtime was kept for 40 min and no peak was observed after 15 min and hence in the subsequent runs we used the run time 15 min. After every 15 min incubation time, sample was injected and the elution peaks obtained were analyzed. Analysis of obtained elution peaks showed a new peak emerging at the retention time ~ 12.5 min (Figures 9C and 9D).

The analysis of elution profile shows that the area and intensity of new peak (~ 12.5 min) were continuously increasing which might be of the free 3mA released from the DNA by the enzymatic action. Analysis was performed till the area and intensity of the new peak

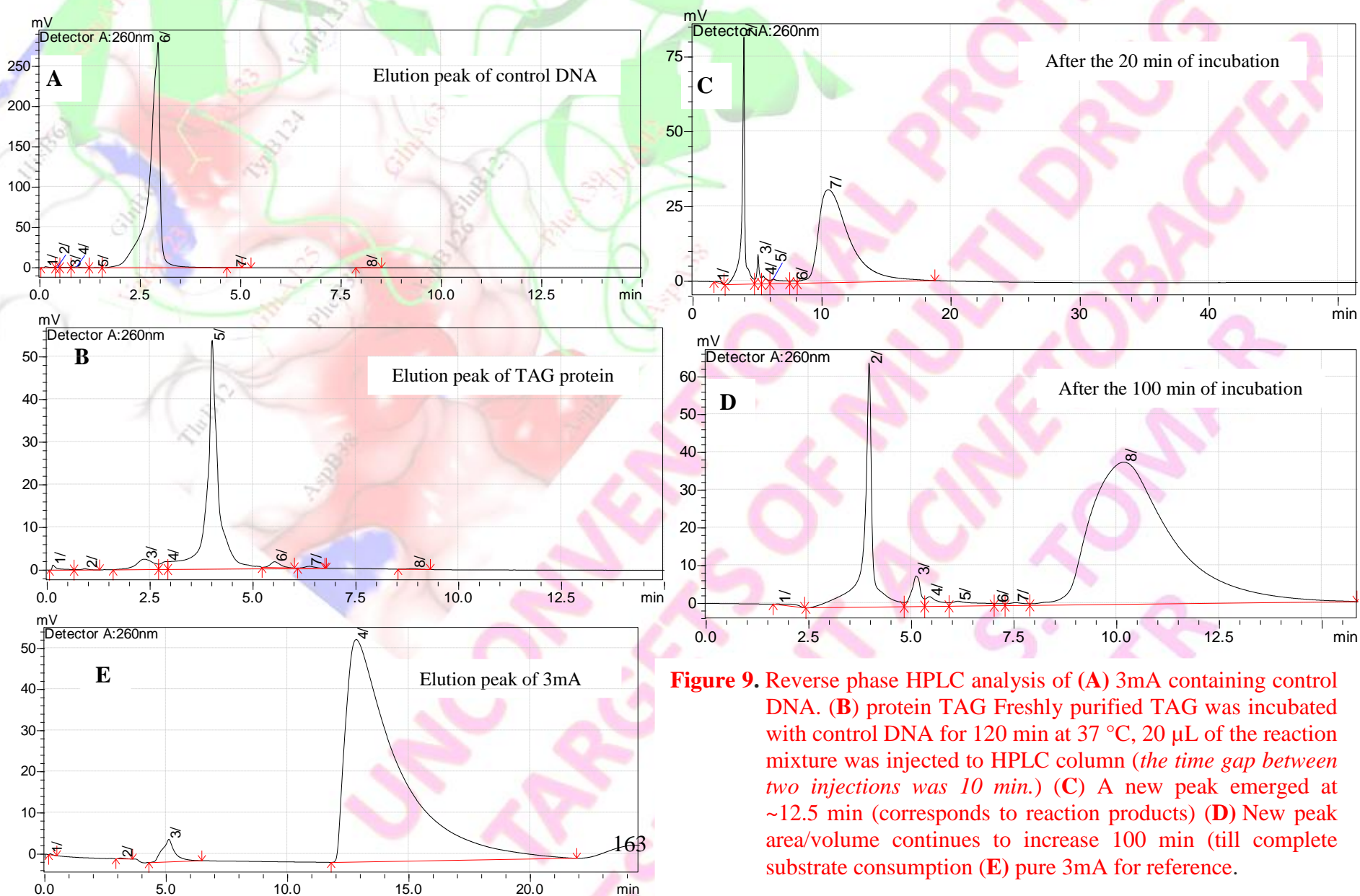


Figure 9. Reverse phase HPLC analysis of (A) 3mA containing control DNA. (B) protein TAG Freshly purified TAG was incubated with control DNA for 120 min at 37 °C, 20 μ L of the reaction mixture was injected to HPLC column (the time gap between two injections was 10 min.) (C) A new peak emerged at ~12.5 min (corresponds to reaction products) (D) New peak area/volume continues to increase 100 min (till complete substrate consumption) (E) pure 3mA for reference.

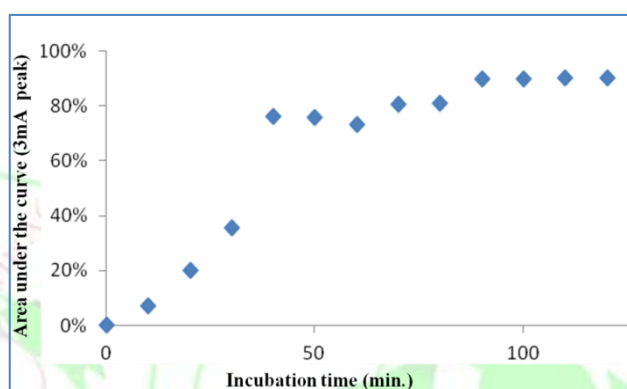


Figure 10. Reversed-phase HPLC of 3-methyl adenine released by TAG from the control DNA at different time interval. The TAG-excised product is eluted at 12.5 min represents similar retention time as obtained for the pure 3mA.

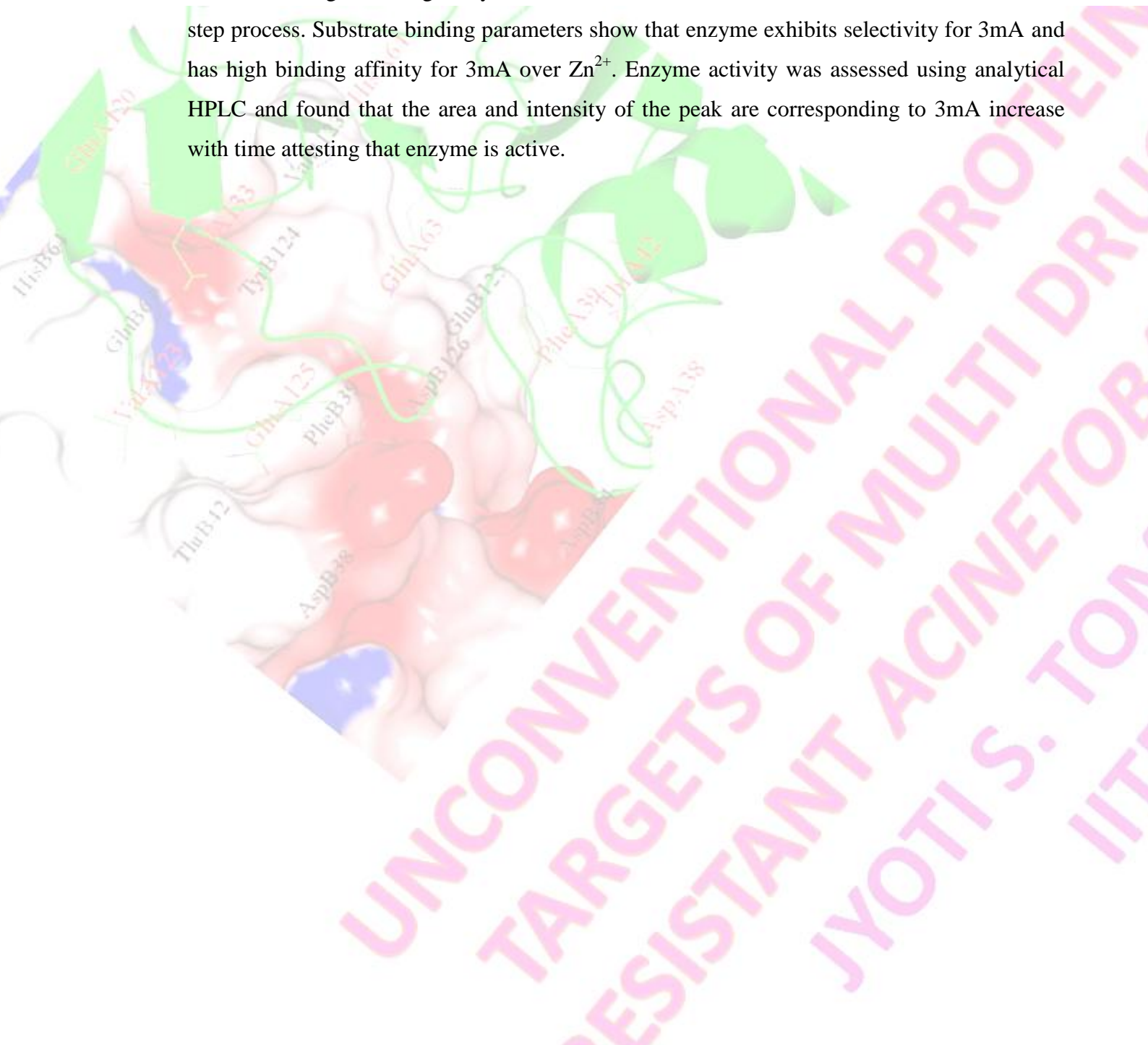
became constant. It was found that intensity and area under the curve of new elution peak were continued to increase till the incubation time 100 min (Figure 10) and later it became constant which means that either entire substrate was consumed or enzyme was completely saturated with substrate (Appendix IV-4). The percentage of the peak area as a function of different incubation time was plotted and analysis showed the increase in the area till 100 min; however, at 100–120 min it showed no variation. The HPLC analysis of the 3mA purchased from the sigma is also performed to obtain the retention time under the same experimental conditions.

The retention time for 3mA is observed here at 12.64 min (Figure 9E). This confirms that the new peak emerging on incubation of TAG with control DNA corresponds to 3mA. The glycosylase assay performed here clearly indicates that the recombinant TAG enzyme is active.

4.5. Conclusion

TAG hydrolyzes the glycosidic bond of methylated adenine in DNA and it was found in various bacteria and some eukaryotes. It exhibits very less primary sequence similarity with members of HhH superfamily of DNA glycosylases. Unlike other glycosylases, it has specificity for 3mA base. In this study, TAG was cloned and expressed in a bacterial expression system. Overexpressed protein was purified using affinity and size exclusion chromatography. However, our initial thoughts were to crystallize the protein for complete

elucidation of its structure. But when those attempts failed, we deduced the structure of TAG using *in silico* modelling. The secondary structure was characterized using CD spectroscopy. Proteins folding/unfolding study was done and results show that TAG denaturation is a two-step process. Substrate binding parameters show that enzyme exhibits selectivity for 3mA and has high binding affinity for 3mA over Zn^{2+} . Enzyme activity was assessed using analytical HPLC and found that the area and intensity of the peak are corresponding to 3mA increase with time attesting that enzyme is active.



4.6. References

1. Brown, S. and Amyes, S. (2006) OXA β -lactamases in *Acinetobacter*: The story so far; J. Antimicrob. Chemother. 57, 1–3.
2. Kaur, N., Khokhar, M., Jain, V., Bharatam, P.V., Sandhir, R. and Tewari, R. (2013) Identification of druggable targets for *Acinetobacter baumannii* via subtractive genomics and plausible inhibitors for MurA and MurB; Appl. Biochem. Biotech. 171, 417–436.
3. Adane, L., Bharatam, P.V., Sharma, V. (2010) A common feature-based 3D-pharmacophore model generation and virtual screening: Identification of potential PfDHFR inhibitors; J. Enzyme Inhib. Med. Chem. 25, 635–645.
4. Thomson, J.M. and Bonomo, R.A. (2005) The threat of antibiotic resistance in gram-negative pathogenic bacteria: β -lactams in peril; Curr. Opin. Microbiol. 8, 518–524.
5. Zhu, J., He F., Hu S., and Yu J. (2008) On the nature of human housekeeping Genes; Trends Genet. 24, 481–484.
6. Wilson, D.M., and Thompson L.H. (1997) Life without DNA repair; Proc. Natl. Acad. Sci. USA 94, 12754–12757.
7. Wilson, M.D., Barbosa-Morais, N.L., Schmidt, D., Conboy, C.M., Vanes, L., Tybulewicz, V.L., Fisher, E.M., Tavaré, S. and Odom, D.T. (2008) Species specific transcription in mice carrying human chromosome 21; Science 322, 434–438.
8. Sartori, A.A., Lukas C., Coates J., Mistrik M., Fu S., Bartek J., Baer R., Lukas J. and Jackson S.P. (2007) Human CtIP promotes DNA end resection; Nature 450, 509–514.
9. De Bont, R. and van Larebeke, N. (2004) Endogenous DNA damage in humans: A review of quantitative data; Mutagenesis 19, 169–185.
10. Lindahl, T. (1993) Instability and decay of the primary structure of DNA; Nature 362, 709–715.
11. Barbin, A. (1999) Role of etheno DNA adducts in carcinogenesis induced by vinyl chloride in rats; IARC Sci. Publ. No. 150. 303–313.
12. Bolt, H.M. (2005) Vinyl chloride – A classical industrial toxicant of new interest; Crit. Rev. Toxicol. 35, 307–323.
13. Voigt, M.D. (2005) Alcohol in hepatocellular cancer; Clin. Liver. Dis. 9, 151–69.

14. Ohtani, N., Mann, D.J. and Hara, E. (2009) Cellular senescence: Its role in tumor suppression and aging; *Cancer Sci.* 100, 792–797.
15. Best, B.P. (2009) Nuclear DNA damage as a direct cause of aging; *Rejuvenation Res.* 12, 199–208.
16. Capell, B.C., Tlougan, B.E. and Orlow, S.J. (2009) From the rarest to the most common: Insights from progeroid syndromes into skin cancer and aging; *J. Invest. Dermatol.* 129, 2340–2350.
17. McCulloch, S.D., Gu, L. and Li, G.M. (2003) Bi-directional processing of DNA loops by mismatch repair-dependent and-independent pathways in human cells; *J. Biol. Chem.* 278, 3891–3896.
18. Gabriele, A. and Margherita, B. (2001) Mismatch repair in correction of replication errors and processing of DNA damage; *J. Cell. Physiol.* 187, 145–154.
19. Inoue, H., Nojima, H. and Okayama, H. (1990) *E. coli* competent cells and transformation; *Gene* 96, 23–28.
20. Bradford, M.M. (1976) Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding; *Anal. Biochem.* 72, 248–254.
21. Bohm, G., Muhr, R. and Jaenicke, R. (1992) Quantitative analysis of protein far UV circular dichroism spectra by neural networks; *Protein Eng.* 5, 191–195.
22. Zitzewitz, J.A., Bilsel, O., Luo, J., Jones, B.E. and Matthews, C.R. (1995) Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy; *Biochemistry* 34, 12812–12855.
23. Greenfield, N.J. (2006) Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism; *Nat. Protoc.* 1, 2733–2741.
24. Kwon, K., Cao, C. and Stivers, J.T. (2003) A novel zinc snap motif conveys structural stability to 3-methyladenine DNA glycosylase I; *J. Biol. Chem.* 278, 19442–19446.
25. Pawlak, S.D., Radlinska, M., Chmiel, A.A., Bujnicki, J.M. and Skowronek, K.J. (2005) Inference of relationships in the ‘twilight zone’ of homology using a combination of bioinformatics and site-directed mutagenesis: A case study of restriction endonucleases Bsp6I and PvuII; *Nuc. Acids Res.* 33, 661–671.

26. Drohat, A.C., Kwon, K., Krosky, D.J. and Stivers, J.T. (2002) 3-Methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member; *Nat. Struct. Biol.* 9, 659–664.
27. Rinne1, M.L., He, Y., Pachkowski, B.F., Nakamura, J. and Kelley, M.R. (2005) N-Methylpurine DNA glycosylase overexpression increases alkylation sensitivity by rapidly removing non-toxic 7-methylguanine adducts; *Nucl. Acids Res.* 33, 2859–2867.
28. Bjelland, S., Birkeland, N.K., Benneche, T., Volden, G. and Seeberg, E. (1994) DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the AlkA enzyme in *Escherichia coli*; *J. Biol. Chem.* 26, 30489-30495.
29. Gogos, A. and Clarke, N.D. (1999) Characterization of an 8-oxoguanine DNA glycosylase from *Methanococcus jannaschii*; *J. Biol. Chem.* 274, 30447–30450.
30. Williams, M.V. and Pollack, J.D. (1990) A mollicute (mycoplasma) DNA repair enzyme: Purification and characterization of uracil-DNA glycosylase; *J. Bacteriol.* 172, 2979–2985.

5.1. Introduction

DNA, the molecular basis of life, has vital importance in understanding the mechanism of cell division, cell differentiation, ageing and senescence. DNA is considered as target for chemotherapy for decades. Generally, it occurs in the form of duplex but it also exists in a variety of secondary structures including triplex [1, 2], cruciform [3], Z-DNA [4] and quadruplex [5–9]. The 3'-terminal extension of G-rich sequence is a conserved feature of human telomeric DNA. Universal conservation of terminal sequence recommends that structures formed by these sequences are significantly involved in telomeric functioning. Variations in telomere status can force the proliferation of both healthy and cancerous cells [10]. The telomeres of somatic cells get shortened [11] due to the end-replication problem. Approximately, 100 bases get lost in every division till the Hayflick limit [12], then cells become senescent [13] or undergo apoptosis.

In 80–85% of tumor cells, the telomere length is maintained by the telomerase enzyme [14, 15]. The telomerase activity is negligible in the normal somatic cells and for that reason it acts as selective target for cancer chemotherapy [16, 17]. The core telomerase components have been targeted directly for chemotherapy for the past decade. But problem associated with this approach was that the capability of cells for adopting telomerase-independent mechanism for telomere maintenance *i.e.*, alternative lengthening of telomere [18]. Direct telomere targeting may also not be safe, as there is no direct evidence supporting the notion that cancer and normal somatic cells have different telomere structures. Thus targeting of G-rich overhang is considered to be more suitable since it is a critical determinant in the formation of higher order telomere structure termed as shelterin complex [19]. Stephen and Gary found that a guanine-rich repeating sequence forms the G-quartet structure [20]. The G-quartet stacks to form G-quadruplex and the formation of G-quadruplex hinders telomerase action as it needs unfolded telomere for elongation [21]. Muniyappa, in 2005, observed that quadruplex structures play some role in chromosome synapsis and recombination during meiosis [22]. Zaug and co-workers observed that hPOT1 disrupts the intramolecular G-quadruplex structure hence targeting hPOT1 and stabilizing G-quadruplex, might have synergistic efficacy [23]. Increasingly reported experimental evidences suggest that interfering with telomeres *via* direct targeting of telomeric G-quadruplex structures, may also be potentially valuable for antitumor therapeutic strategy [24]. Furthermore, it was found that quadruplex-stabilizing ligands represent high correlation between quadruplex affinity and telomerase inhibition [25, 26]. Most of the tested quadruplex-stabilizing ligands are polyaromatic molecules that bind to end of the G-quadruplex, but none of these compounds passed the clinical trials because of their poor drug-like properties. One of the most common drawbacks noticed in these compounds are the poor solubility due to which these compounds were discarded during ADME test. Screening of the dietary compounds having capacity to stabilize G-quadruplex might be one of the possible solutions to circumvent the problems associated with the solubility.

The interesting thing we noticed in the literature study that Arora and co-workers classified different G-quadruplex binding molecular scaffolds and they kept flavonoids in the class of ligands with protonable side arms [27]. Flavonoids are dietary compounds which belong to a group of polyphenolic compounds present in green vegetables, fruits and beverages, and they belong to one of the most studied classes of plant metabolites. Structurally, they consist of two main groups, the 2-arylchromans and the 3-arylchromans. Furthermore, the literature data reveals

that isoflavones possess anticancerous activity along with differential action on G-quadruplex versus duplex and belongs to 3-arylchromans group of flavonoids. According to Zhang and co-workers, genistein, daidzein and their glycosides exhibit the distinctive regulation on structure competition of duplex versus quadruplex [28]. Jin, through experimental study, concluded that daidzein, quercetin, luteolin and berberine have capacity to stimulate the telomeric DNA to form G-quadruplex structure [29, 30]. Isoflavones not only exhibit the capacity of induction of the G-quadruplex formation, but also show tendency of modifying or imposing conformational changes in the G-quadruplex present in the cell. Isoflavones mainly occur as aglycone, glycosidic conjugates or as prenylated isoflavones. Among all, daidzein and genistein are widely reviewed. Isoflavones exhibit anticancerous activity against colon [31], breast [31–34], prostate [35], cervix [36], lung [37], and brain [38] cancers and the data obtained from *in vitro* and *in vivo* studies indicate the capacity of isoflavones in cancer chemoprevention. Till date, various mechanistic modes of action of isoflavones have been identified, including cell cycle arrest [39], induction of apoptosis, prevention of oxidative damage [40, 41], reduction of angiogenesis [42] reduce cell proliferation [43] and inhibition of telomerase activity [44, 45]. However, main cellular target of isoflavones yet to be identified. Inspection of literature data suggests that the induction of a quadruplex formation or change of a quadruplex conformation owing to binding is a powerful method to impose desired biological effect, *i.e.*, anti-cancerous action. Hence it is imperative to perform the binding study of isoflavones, *O*-methylated isoflavones, prenylated isoflavones, and their glycosides with G-quadruplex and duplex DNA. The experimental study is difficult to perform for the large search space and in such cases the theoretical techniques emerge as best option to minimize the search space. The aim of present work is to identify the pioneering isoflavones molecular scaffolds exhibiting high binding affinity and selectivity for the G-quadruplex, having suitable ADME profile and other desired characteristics to enter into higher phases of drug designing. Hence molecular docking is performed in this study as the preliminary step for elucidating the isoflavones action on G-quadruplex. In docking simulations, the isoflavones explore all the available binding sites on the G-quadruplex and duplex DNA structure and then free-energy changes of the whole binding process is computed using the GBSA. We found a peculiar hopping groove binding mechanism for the isoflavones. The selectivity and affinity of the selected scaffold can be improved further by structural modifications, as availability of enormous structural diversity provides an opportunity to investigate all different combinations of functionalized side chains. These findings suggest the scope for *in vitro* and *in vivo* studies to be done for the validation of theoretical results.

5.2. Materials and Methods

The parallel G-quadruplex d-(TTAGGGT)₄ structure was retrieved from the protein data bank (PDB:ID-1NP9) (<http://www.rcsb.org/pdb>). The duplex B-DNA d-(TTAAGGGT)₂ model was created using *Build tool* of MOE [46]. Prior to theoretical calculations, the DNA structures were subjected for minimization with rms gradient of 0.01 kcal Å⁻¹ mol⁻¹ and AMBER99 force-field in MOE. Energy minimization in MOE was done by applying non-linear optimization techniques. MOE uses success of three methods–Steepest descent, Conjugate gradient and Truncated Newton. The Steepest descent method is very fast and stable initially but extremely inefficient after few iterations

due to very slow convergence. The Conjugate gradient method improves upon-Steepest descent by choosing the next search direction keeping the progress accomplished from the previous step. The 3D structures of isoflavones were obtained from NCBI (<http://www.ncbi.nlm.nih.gov/pccompound>). These structures were optimized using MMFF94 force-field in MOE. Structures of isoflavones used in this study are given in **Figure 1**. These structures were used as initial coordinates for the docking study.

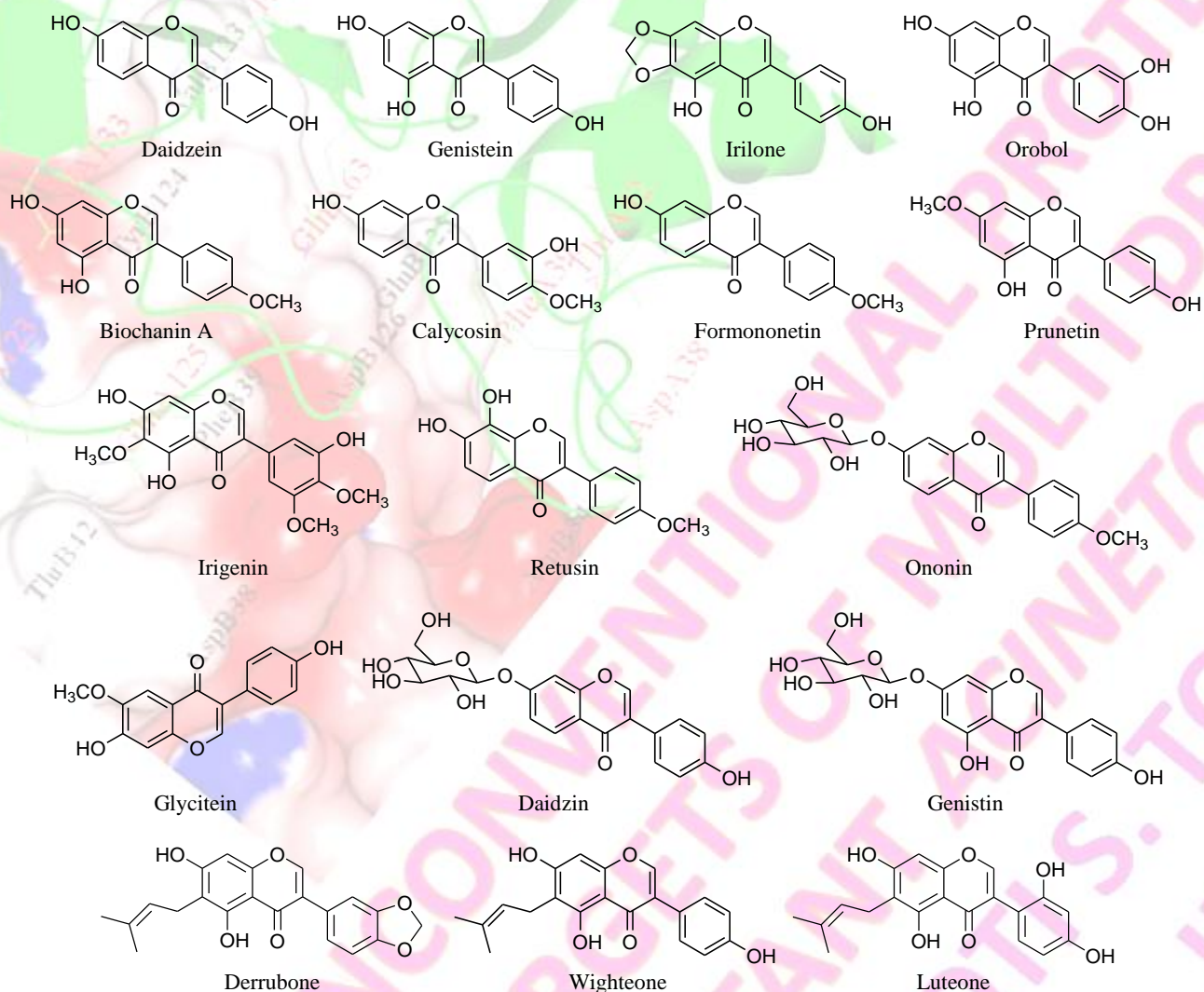


Figure 1. Isoflavones taken for molecular docking simulation.

5.2.1 Validation of docking software and parameters

Validation of the docking technique is necessary prior to the screening of a particular molecular system to validate the software and docking parameters. In this study, redocking experiment was performed to examine whether molecular docking simulations using Autodock4 are able to reproduce the binding pose as reported in experimental G-quadruplex complexes or not. In redocking (bound docking), the atomic structures from a co-crystal are used as

starting coordinates while in unbound docking, the initial structures are native means unbound. For the redocking, almost all types of G-quadruplex ligand complexes, present in the pdb, were downloaded, minimized and then subjected for the docking run. After completion of docking run, the best theoretical binding poses of each experimental ligand were saved. The rmsd values for the theoretical pose were calculated keeping the experimental pose as reference.

5.2.2 Molecular Docking studies of isoflavones

The ability of docking to predict physical configuration, which can optimize favorable interactions is eminently valuable in the drug designing. Molecular docking is performed using Autodock (version 4.0.2) [47] and Dock6.4 [48] modeling packages. For exploration of conformations, Lamarckian genetic algorithm was preferred. LGA is a hybrid of genetic algorithm (for global search) and Solis and Wets algorithm (for local search). In LGA, the ligand state is represented by chromosome whereas a gene, unit of chromosome symbolizes the co-ordinates, orientation and conformation of ligands. The simulation initiates by creating a random population and best intermediate chromosome is subjected to mutation and crossover to produce the next generation. Each generation cycle is followed by a local search. The best solutions were scored by force-field method which evaluates the binding in two steps. In the first step, it calculates the intramolecular energetics of transition from unbound to bound conformation and in the second step, it evaluates the intermolecular energetics of complex. In Autodock4, force-field includes conformational entropy loss (ΔS_{conf}) along with pairwise energy term. Autodock performs rapid energy evaluations through pre-calculated grids of affinity potentials. The dimension of grid box was sufficiently large to accommodate whole molecule inside. Thus, grid points of $90 \times 90 \times 90$ with a 0.375 \AA spacing were calculated around docking area for all the isoflavones using AutoGrid4.2. Default parameters were used with 5,000,000 numbers of energy evaluation and rmsd value of 2.5.

Dock6.4 is one of the best known molecular modeling packages consisting of two key steps: a search algorithm and scoring function. The search algorithm samples the orientations of receptor and ligand. Anchor-and-grow algorithm allows optimization of ligand orientation according to the shape of active site. Scoring function ranks the geometry generated by search algorithm. The generalised Born and surface area solvation score (GBSA score) in dock6.4 is an implementation of molecular mechanics with GBSA (MM/GBSA), which uses pairwise GB salvation model to compute free energies for binding of ligands. The employment of GBSA evidently improves the screening and ranking power and is reviewed by many researchers [49, 50]. Generally GBSA score is used as the second step screening after a prescreening by the force-field method as GBSA rescoring performs better than the normal docking score. Hence in this study, the calculation of GBSA score was performed for an accurate estimation of binding free energies and nature of molecular interactions involved in the complex formation. These calculations allowed us to identify the binding sites of isoflavones and revealing the lowest energy ligand binding modes to better discriminate the true hit from the false hit. Furthermore, it is useful in estimating the free energy changes of solute-solvent interactions during various bio-chemical processes and provides valuable information indispensable for drug designing.

5.3. Results and discussion

5.3.1 Validation of results obtained through redocking experiment

The experimental G-quadruplex complexes are reproduced for the validation step. The redocked ligands poses superimposed on the corresponding experimental poses are given in **Figure 2**. For the rmsd calculation, we have considered not only the highest ranking conformation, but also the conformation having smallest rmsd value. The less significant rmsd values obtained between redocked and experimental conformations suggest that the performance of Autodock is satisfactory. Analysis of the obtained redocking results indicate that 1:1 G-quadruplex complexes were successfully reproduced by the docking program, while in 1:2 G-quadruplex complexes, the ligands were slightly biased toward the first binding site as a consequence of additional favorable interactions at the first binding site. Also,

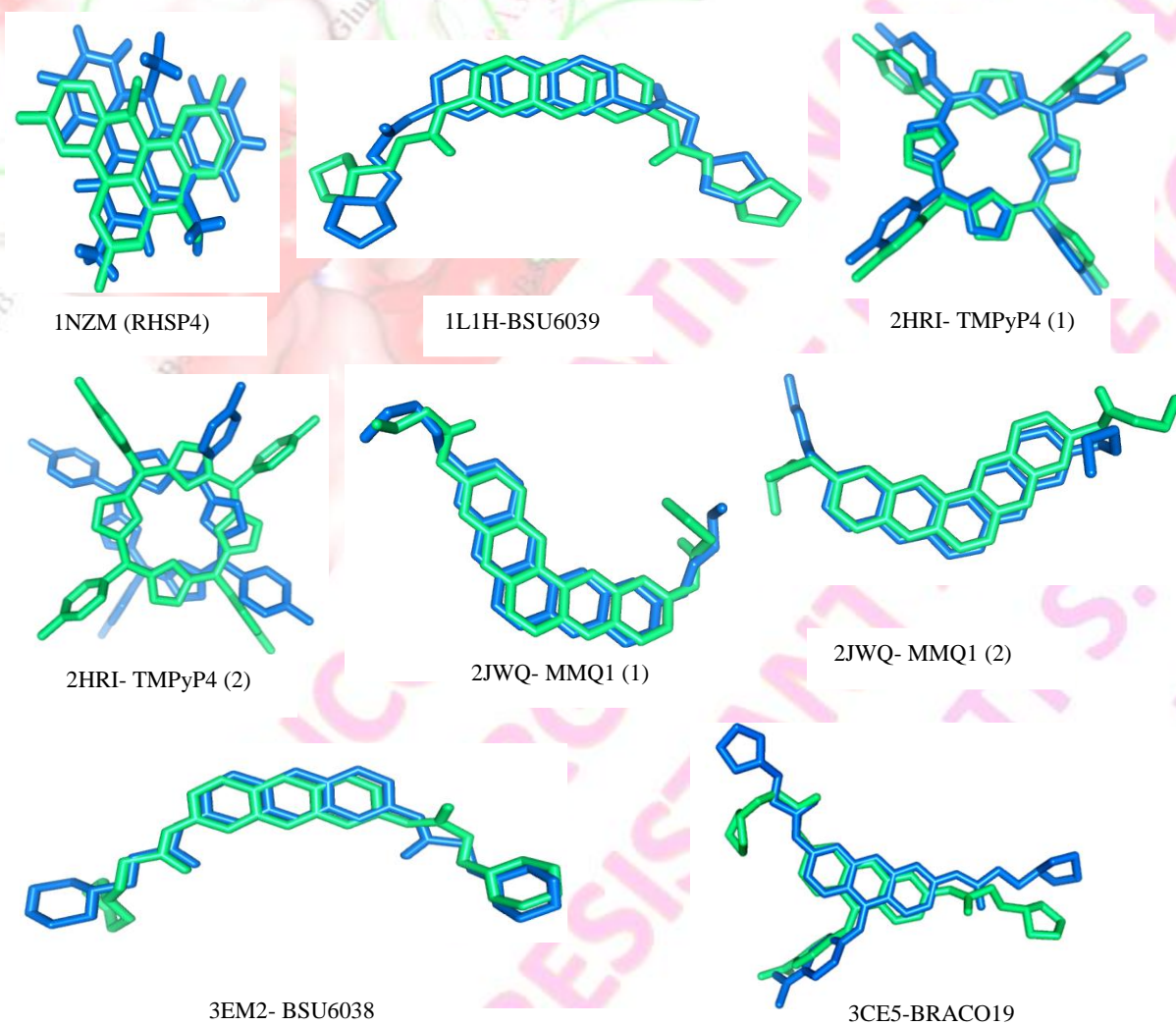
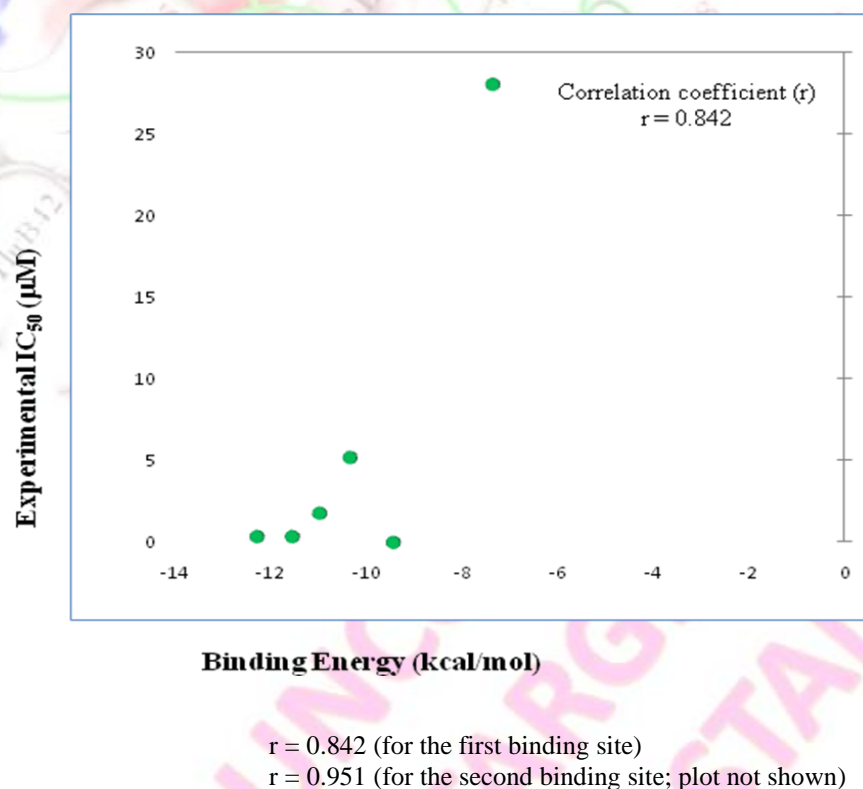


Figure 2. Docked pose (Green) is superimposed on the experimental pose (Blue). Compound name and PDBid is given along with the respective image.

in most of the complexes, the first binding site is deeper and contributes to the stability by minimizing the unfavorable interactions between hydrophobic core of ligand and water. Due to these reasons, the second binding site is calculated by specifying the grid box by excluding the first binding site. Recent studies on validation of docking programs for ligands–DNA docking process by Holt and co-workers [51] and Netz and co-workers [52] have shown that Autodock can accurately reproduce the experimental complex structures. The employment of docking techniques to investigate the binding mechanism of new compounds with the DNA is a challenging task. The DNA ligand binding process consists of three styles *viz.*, stacking, groove binding and intercalation. The first two binding modes do not impose any structural distortion to the target and the ligand binding process is comparable with the lock and key model of enzyme action. But in the third binding mode, the formation of intercalation cavity compels DNA to undergo structural rearrangement like induced fit model of enzyme action, due to which the prediction of the correct intercalation pose is very tedious and still challenging for most of the docking softwares. To handle such situations, Netz and co-workers suggested that the employment of docking technique without any prior experimental proof of binding mode needs canonical DNA with and without an intercalation cavity to investigate the correct binding mode



[52]. Theoretical binding energies, rmsd values and IC₅₀ values for all the redocked ligands are given in Table 1. The Pearson's correlation coefficient (r = 0.842) is calculated which indicates statistically significant correlation between theoretical binding energies vs. experimental IC₅₀ values and the plot is shown in Figure 3. This further reveals that the docking simulation is successful in discriminating various ligand G-quadruplex complexes.

Figure 3. Binding energy vs. experimental IC₅₀ (µM) plot for experimental G-quadruplex complexes.

PDBid (LIGAND)	RMSD Value	Binding Energy (kcal/mol)	Experimental IC ₅₀ Value	DNA Sequence	Reference(s)
1NZM (2:1) (RHSP4)	0.23 0.98	-11.53 ± 0.13 -10.63 ± 0.20	0.330 μM	d(TTAGGGT)	[59]
2JWQ (2:1) (MMQs)	1.01 1.56	-12.24 ± 0.17 -10.47 ± 0.14	0.373 μM	d(TTAGGGT)	[60]
2HRI (2:1) (TMPyP4)	0.89 3.62	-07.35 ± 0.21 -05.01 ± 0.23	28.00 μM	d(TAGGGTT AGGG)	[61,62]
3EM2 (1:1) (BSU-6038)	1.32	-10.95 ± 0.15	~1.800 μM	d(GGGGTTT TGGGG)	[60]
1L1H (1:1) (BSU6039)	1.62	-10.33 ± 0.19	~5.200 μM	d(GGGGTTT TGGGG)	[60]
3CE5 (1:1) (BRACO-19)	3.56	-09.44 ± 0.11	0.020 μM	d(GGGGTTT TGGGG)	[60]

Table 1. G-Quadruplex complexes: Comparison of binding energies and rmsd values with experimental IC₅₀ values.

5.3.2 Analysis and comparison of binding interactions in experimental and redocked ligand poses of ligand-G-quadruplex complexes

The rmsd values are extensively acceptable and reliable to decide the docking accuracy perhaps may not be ideal but are used by majority of researchers. The deviation in rmsd values for the redocked conformation is found to be roughly proportional to the number of torsions present. For example, the rmsd value is found to be lowest for the RHSP4 ligand having no rotatable bonds, while the rmsd value is highest for BRACO-19 having maximum number of rotatable bonds (11 bonds). The presence of many rotatable bonds in ligand makes it more vulnerable for the wrong pose prediction and with increase in the number of torsion angles, the conformational space increases exponentially and pose prediction process becomes more complex. The analyses of binding sites are done to examine the molecular interactions involved in the binding. BRACO-19, a tri-substituted acridine derivative, binds at the 3' end of G-quadruplex and is stabilized by the π - π interaction with two guanine residues. Its 3,6-bisamido side chain substituent extends toward wide groove and 9-anilino substituent fits into narrow groove to maximize the binding interactions. While in redocked conformation, the positions of both side chain substituents are slightly different from the experimental conformation; 9-anilino substituent flipped toward the wider groove and oxygen atoms of 3,6-bisamido group are involved in the direct hydrogen bonding with G5 and G11 residues. The ligand di-substituted acridine derivative BSU-6039 is sandwiched between G/T base pair and is stabilized by the stacking interactions. The slight change in the conformation of side chains of redocked ligand neglects the possibility of hydrogen bond formation present in the experimental pose. The RHSP4 ligand binds at the ends and is stabilized by the stacking interactions with G4 and G5 residues of parallel G-quadruplex. Docked and experimental poses are almost identical for RHSP4. The MMQ1 ligand also binds at the ends of parallel G-quadruplex, and is stabilized by the stacking

interactions. Both RHSP4 and MMQ1 ligands consist of equal number of aromatic rings but possess quite different scaffolds and binding mode. Of the stacking interactions present in MMQ1 and RHSP4 with the ligand, the interactions are more effective in the former owing to its crescent shape. The two molecules of TMPyP4 ligands bind onto the TTA nucleotides by the stacking interaction instead of guanine nucleotide, *via* external loop structure and the 5' region of the stacked quadruplex and thus no direct interactions between TMPyP4 and G-tetrads may be seen. The docked pose of TMPyP4 forms a hydrogen bond with T6:H3, which is not present in experimental complex of TMPyP4.

5.3.3. Analysis of results from molecular docking of isoflavones

Molecular docking for nucleic acids using small molecules is still at its infancy. To overcome the problems associated with DNA docking, one should have some prior experimental knowledge or may follow the strategy described by the Netz and co-workers [52]. Information obtained from the experimental and computational studies performed by many research groups suggests that isoflavones possess groove binding or stacking interactions. For example, binding of daidzein with intramolecular G-quadruplex was studied by Li using CD, ESI-MS, PAGE and molecular simulations and concluded that the complex is stabilized by π - π and hydrogen bonding interactions [29]. Similarly, Zhang and co-workers found that isoflavones prefer to bind at G-quartet plane, diagonal loop and groove [28]. Based on the experimental information available about binding mode of the isoflavones, the requirement of DNA molecule along with the intercalation cavity can be neglected. Then the docking simulations of isoflavones with G-quadruplex and duplex DNA were performed. The docking analysis of isoflavones with quadruplex and duplex reveals that the isoflavones possess preferential binding with G-quadruplex as compared to duplex. Binding energies, inhibition constant values of isoflavones with duplex are given in Table 2. In recent multistep virtual screening experiments by the Huang and co-workers, it is concluded that if $\text{Score-G4}/\text{Score-groovebinder} > 1.1$, the compounds are selected for visual inspection and experimental validation [53]. In the current study, we have also followed the same strategy to evaluate our docking results for the selectivity, and we found that prenylated isoflavones have better selectivity followed by the glycosidic derivatives, while the remaining isoflavones except orobol, biochanin A, calycosin and glycitein do not possess any selectivity for quadruplex over duplex DNA. The G-quadruplex-isoflavone complexes are subjected to X3 DNA/RNA 2.0 program [54] to find out whether any structural changes take place during the complexation process. Obtained X3 DNA results for the structural parameters indicate that the G-quadruplex structure remains nearly intact during the formation of complex indicating the lock and key binding accomplishment of isoflavones. Binding energies, inhibition constants, and target residues involved in the hydrogen bonding of best docked poses are given in Table 3. Each G-quadruplex displays four grooves and the dimensions of these grooves depend on the direction of strand alignment and topologies of the connecting loops [55]. Parallel G-quadruplex has four equivalent grooves. The nature, as well as dimensions of duplex and G-quadruplex grooves is very different, which implies that the targeting of G-quadruplex groove might aid to the selectivity of G-quadruplex upon duplex. The results from the docking simulation show that isoflavones possess more than one binding site because of the equivalent nature of all four grooves. Daidzein, genistein and formononetin molecules

bind in the groove present between the chains A and D. The methylated isoflavones, except formononetin, bind at the groove between the chains B and C near to guanine bases. All of these compounds are held inside the groove hydrogen bonding and van der Waals interactions. The values of calculated binding energy differentiate the binding potential of genistein and daidzein from their glycosides, similar to the conclusion drawn by Zhang and co-workers [28]. The stronger binding of glycosides as compared to aglycone may be attributed to the propensity of carbohydrate moiety toward the formation of several hydrogen bonds. The potential sites for hydrogen bonding in groove walls are *N2* amino and *N3* atoms of guanine bases, *N3* atom of adenine, *O4* atom of sugar and phosphate backbone atoms. The presence of potential sites for hydrogen bonding in substrate and receptor helps in the formation of more number of hydrogen bonds and provides significant contribution in stabilization of ligands. The results from ESI-MS show that the amount of daidzein and genistein bound to per G-quartet are 0.90 and 0.86 mM, respectively [28]. This illustrates that daidzein possesses higher binding affinity than genistein; similarly in the present study daidzein was found to be having more binding affinity than genistein.

<i>Isoflavones</i>	<i>No. of poses in optimum cluster (out of 10)</i>	<i>Binding energy (kcal/mol)</i>	<i>Inhibition constant (Ki)</i>
Daidzein	6	-6.13 ± 0.07	23.7 μM
Genistein	6	-6.01 ± 0.19	39.2 μM
Irilone	6	-6.52 ± 0.20	16.1 μM
Orobol	4	-5.45 ± 0.10	115.7 μM
<i>O-methylated isoflavones</i>			
Biochanin A	6	-5.38 ± 0.13	59.1 μM
Calycosin	9	-5.10 ± 0.11	142.2 μM
Formononetin	4	-6.40 ± 0.07	18.4 μM
Glycitein	4	-5.07 ± 0.12	176.6 μM
Irogenin	8	-6.13 ± 0.15	29.2 μM
Prunetin	8	-6.26 ± 0.09	31.4 μM
Retusin	6	-6.06 ± 0.11	25.3 μM
<i>Isoflavones Glycosides</i>			
Daidzin	5	-5.05 ± 0.10	198.1 μM
Genistin	4	-4.46 ± 0.20	461.3 μM
Ononin	5	-5.21 ± 0.07	153.7 μM
<i>Prenylated isoflavones</i>			
Derrubone	5	-5.11 ± 0.10	179.7 μM
Luteone	4	-5.26 ± 0.09	149.5 μM
Wighteone	4	-4.75 ± 0.11	361.4 μM

Table 2. Binding data of isoflavones with TTAGGGT duplex

Prenylation of isoflavones makes these compounds more active and it increases the lipophilicity and thereby membrane permeability [56]. Generally, prenylated compounds are considered as natural medicine and lead compounds for drug designing [57]. Analysis of various binding parameters symbolizes that isoflavones are most potent binders and have better selectivity for the quadruplex groove binding mode than other derivatives of isoflavones. Binding site and binding conformations of derrubone and luteone are very similar *i.e.*, groove formed by the chains A and D near to the residues A3G4G5. In particular, four hydrogen bonds are formed between the 7-OH, 4-carbonyl oxygen and benzodioxol moiety of the derrubone and G5G4A3 bases of DNA. The binding position of wighteone is different and it binds to

<i>Isoflavones</i>	<i>No. of poses in optimum cluster (out of 10)</i>	<i>Binding energy (kcal/mol)</i>	<i>Inhibition constant (Ki)</i>	<i>Residues involved in H-bonding</i>	<i>Score-G4/Score-Groovebinder >1.1</i>
Daidzein	8	-6.63 ± 0.02	13.7 µM	A3,G4,G5 (3-H bonds)	1.08
Genistein	8	-6.38 ± 0.08	21.2 µM	T2,A3,G4, G25 (4-H bonds)	1.06
Irilone	6	-7.52 ± 0.10	03.1 µM	A17,G18,G19 (3-H bonds)	1.15
Orobol	4	-7.15 ± 0.19	05.7 µM	A10,A17,G18 (4-H bonds)	1.31
<i>O-Methylated isoflavones</i>					
Biochanin A	4	-6.88 ± 0.15	09.1 µM	G18,G26 (3-H bonds)	1.26
Calycosin	5	-7.10 ± 0.11	06.2 µM	A10,A17 (3-H bonds)	1.39
Formononetin	8	-6.80 ± 0.03	10.4 µM	G4,G5,A24 (3-H bonds)	1.06
Glycitein	6	-7.07 ± 0.12	06.6 µM	A10,A17,G18 (3-H bonds)	1.39
Irogenin	4	-6.43 ± 0.23	19.2 µM	G4 (1-H bond)	1.05
Prunetin	5	-6.56 ± 0.15	15.4 µM	A17,G18,G19 (3-H bonds)	1.05
Retusin	4	-6.26 ± 0.09	25.3 µM	G25,G26,G27 (3-H bonds)	1.03
<i>Isoflavonyl glycosides</i>					
Daidzin	8	-7.77 ± 0.14	02.1 µM	A24,G25,G26, G27,G18 (5-H bonds)	1.54
Genistin	7	-6.46 ± 0.14	18.3 µM	T8,A17,G18 (4-H bonds)	1.45
Ononin	8	-7.41 ± 0.16	03.7 µM	A3,G4,A24 (3-H bonds)	1.42
<i>Prenylated isoflavones</i>					
Derrubone	6	-8.38 ± 0.11	00.7 µM	G4,G25,G26 (4-H bonds)	1.64
Luteone	6	-8.56 ± 0.09	00.5 µM	G18,G19,G26, G27 (5-H bond)	1.63
Wighteone	5	-8.75 ± 0.19	00.4 µM	A10,A17,G18 (3-H bond)	1.84

Table 3. Binding energy, inhibition constants and residue numbers involved in the H-bonding between G-quadruplex and isoflavones.

groove formed by the chains B and C near residues to A10A17G18. Binding affinities of prenylated compounds are comparable to well known quadruplex stabilizer and telomerase inhibitor TmPyp4. These prenylated isoflavones can be considered as better lead candidates for further studies as these are natural dietary products and do not have toxicity and solubility problems.

5.3.4. Docking performance improvement using GBSA rescoring

Molecular mechanics/generalized born surface area (MM/GBSA) calculations using dock6.4 were performed for the best ranking conformation molecules to throw some light on the solvation forces involved in the stabilization of complexes. In GBSA, the term GB represents the electrostatic component of salvation, while SA represents non- electrostatic contribution termed as cavitation energy. MM/GBSA is one of the most common methods among the implicit solvent models. The recent study on G-quadruplex DNA groove binding molecules suggests, that flexibility of DNA plays a vital role during the ligand binding and water molecules contribute majorly in stabilizing the most relevant poses of ligands [58]. The solvation energy terms involved in the formation of complexes are given in Table 4. Analysis of the values of these energy terms illustrates that the van der Waals interactions play a pivotal role in holding of these compounds inside the groove. The negative value of van der Waals interactions reveals good shape complementarities and packing effects on complexation. ES term includes the interaction between charges, hydrogen

<i>Isoflavones</i>	<i>van der Waals</i>	<i>SA</i>	<i>ES-GB</i>
Daidzein	-25.63	-3.53	+4.1
Genistein	-25.43	-3.48	+5.8
Irilone	-27.91	-3.61	+7.1
Orobol	-26.38	-3.78	+11.4
<i>O-Methylated isoflavones</i>			
Biochanin A	-14.64	-2.95	+3.5
Calycosin	-11.54	-2.93	+3.1
Formononetin	-27.32	-3.84	+7.0
Glycitein	-25.37	-3.55	+1.5
Irogenin	-15.24	-3.01	+1.9
Prunetin	-19.25	-3.23	+1.8
Retusin	-13.88	-2.95	+5.3
<i>Isoflavones glycosides</i>			
Daidzin	-38.04	-4.76	+15.1
Genistin	-37.49	-4.97	+15.7
Ononin	-36.31	-4.74	+7.5
<i>Prenylated isoflavones</i>			
Derrubone	-36.71	-4.69	-4.2
Luteone	-34.91	-4.50	+6.4
Wighteone	-32.77	-4.39	+5.5

Table 4. Energy terms vdW, ES and ES-GB between G-quadruplex and isoflavones are calculated using GBSA method.

bonds and electrostatics of desolvation during the binding of isoflavones. The GB term represents change in shape and size of solvent cavity arising from the hydrophobic effects taking place during complexation. In the binding of isoflavones the unfavorable contribution coming in the form of GB term is nearly neutralized by the ES term but still cumulative effect of these two forces have the destabilizing effect for all the isoflavones except the derrubone. In the case of derrubone, the presence of benzodioxol moiety is supposed to be responsible for the stabilizing effect attesting to its optimistic binding mode. van der Waals interactions are found to be the main contributor to the stabilizing term for the isoflavones as well as to the overall interaction energy. The overall free energy change for prenylated isoflavones binding is more due to the well compensated energy terms. Derrubone-G-quadruplex interaction is favored by all the energy terms involved in the binding which signifies its more optimal and strong binding in comparison to other isoflavones (Figures 4).

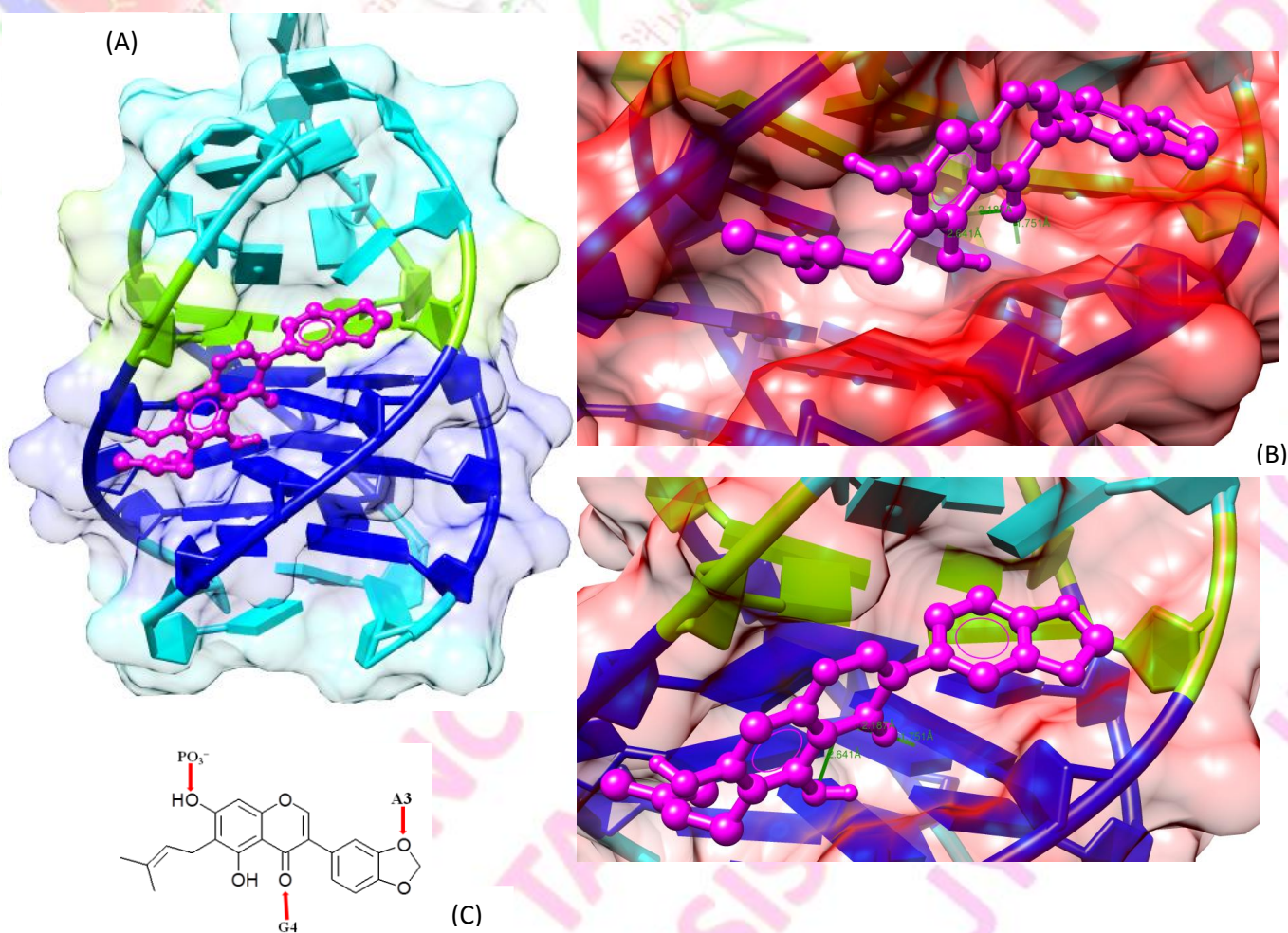


Figure 4. Binding interaction of derrubone with G-quadruplex; (A) Complete view. (B) Closer view (C) H-bonding sites between derrubone and G-quadruplex.

5.4. Conclusion

The initial goal of this work was to determine whether the docking simulation using software Autodock can be used for docking of molecular systems, *i.e.*, G-quadruplex complexes. This was accomplished by selecting the ligand from experimental G-quadruplex complexes. The docking software was found to be able to successfully reproduce the experimental complexes to a high degree of accuracy. The next facet of this work was the empirical validation of result from *in silico* redocking using the experimental IC₅₀ values which exhibited significant correlation with theoretical binding energies. Herein, natural dietary compounds isoflavones were taken for the theoretical study to examine their stabilizing effect on G-quadruplex structure. The experimental G-quadruplex complexes were reproduced to obtain and validate the theoretical parameters. The obtained theoretical binding energies are in significant correlation with the experimental data. Analysis of binding shows isoflavones to be groove binders, and differential nature of quadruplex grooves might be beneficial in the selectivity aspects. Among all, derrubone was found to have better selectivity as well as affinity for the G-quadruplex comparable to well known ligand TMPyP4. The GBSA rescoring result enlightens various interaction terms involved in the binding process. Cumulative stabilizing effects coming from vdW, ES and GB energy terms attest to optimal binding of derrubone molecule which can be considered as a lead for the higher phases of drug designing. Binding poses and energetics for G-quadruplex isoflavones complexes were calculated using molecular docking afterword, rescoring using GBSA was performed. The results corroborated that prenylated and glycosidic isoflavones significantly stabilize the G-quadruplex. The binding conformation and binding energetics of derrubone is most optimal in comparison to other isoflavones. Present study yielded derrubone, from subset of isoflavones, having high binding affinity as well as selectivity for G-quadruplex with theoretical binding efficiency in nanomolar range. Finally, we propose the *in vivo* testing of derrubone which may lead to the discovery of novel natural molecules as lead compounds having desired anticancerous activity. These findings are of great value in terms of unexplored groove binding modes and the studied natural compounds might be helpful to direct the focus of synthetic chemists in designing of new generation of antitumor agents.

5.5. References

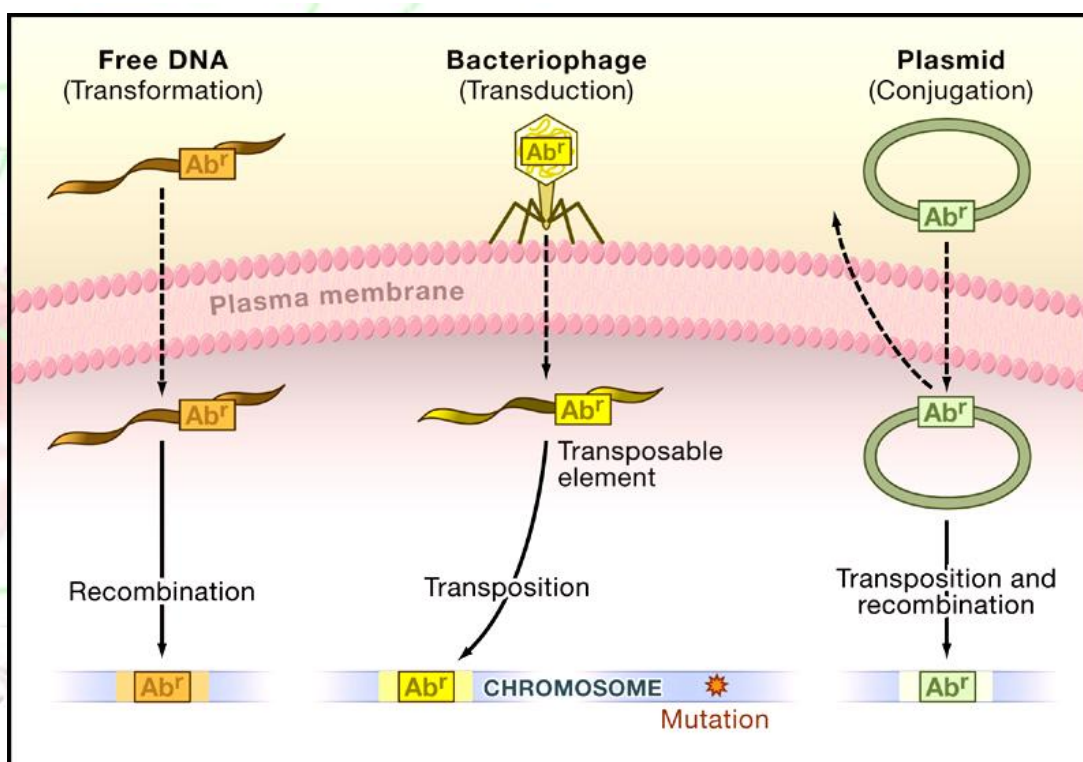
1. Moser, H., Dervan, P. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation; *Science* 238, 456–450.
2. Cooney, M., Czernuszewicz, G., Postel, E., Flint, S. and Hogan, M. (1988) Site-specific oligonucleotide binding represses transcription of the human *c-myc* gene *in vitro*; *Science* 24, 456–459.
3. Timsit, Y. and Moras, D. (1996) Cruciform structures and functions; *Q. Rev. Biophys.* 29, 279–307.
4. Wang, G. and Vasquez, K. (2007) Z-DNA, an active element in the genome; *Front. Biosci.* 12, 4424–4438.
5. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure; *Nucl. Acids Res.* 34, 5402–5415.
6. Han, H. and Hurley, L. (2000) G-quadruplex DNA: A potential target for anti-cancer drug design; *Trends Pharmacol. Sci.* 21, 136–142.
7. Eddy, J. and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome; *Nucl. Acids Res.* 34, 3887–3896.

8. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome; *Nucl. Acids Res.* 33, 2908–2916.
9. Todd, A.K., Johnston, M. and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA; *Nucl. Acids Res.* 33, 2901–2907.
10. Blackburn, E.H. (2000) Telomere states and cell fates; *Nature* 408, 53–56.
11. Harley, C.B., Futcher, A.B. and Greider, C.W. (1990) Telomeres shorten during ageing of human fibroblasts; *Nature* 345, 458–460.
12. Hayflick, L. (1965) The limited *in vitro* life time of human diploid cell strains; *Exp. Cell Res.* 37, 614–636.
13. Shay, J.W. and Wright, W.E. (2000) Hayflick limit and cellular ageing; *Nat. Rev. Mol. Cell Biol.* 1, 72–76.
14. Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer; *Science* 266, 2011–2015.
15. Neidle, S. and Parkinson, G.N. (2002) Telomere maintenance as a target for anticancer drug discovery; *Nat. Rev. Drug Discov.* 1, 383–393.
16. White, L.K., Wright, W.E. and Shay, J.W. (2001) Telomerase inhibitors; *Trends Biotechnol.* 19, 114–120.
17. Mergny, J.L., Riou, J.F., Mailliet, P., Teulade-Fichou, M.P. and Gilson, E. (2002) Natural and pharmacological regulation of telomerase; *Nucl. Acids Res.* 3, 839–865.
18. Lundblade, V. (2002) Telomere maintenance without telomerase; *Oncogene* 21, 522–531.
19. Titia, de. L. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres; *Genes Dev.* 19, 2100–2110.
20. Stephen, N. and Gary, P. (2003) The structure of telomeric DNA; *Curr. Opin. Chem. Biol.* 13, 275–283.
21. Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) Inhibition of telomerase by G-quartet DNA structures; *Nature* 350, 718–720.
22. Muniyappa, K., Anuradha, S. and Byers, B. (2000) Yeast Meiosis-Specific Protein Hop1 Binds to G4 DNA and Promotes Its Formation; *Mol. Cell Biol.* 20, 1361–1369.
23. Zaug, A.J., Podell, E.R., and Cech, T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension *in vitro*; *Proc. Natl. Acad. Sci. USA* 102, 10864–10869.
24. Mergny J.L. and Claude, H. (1998) G-quadruplex DNA: A target for drug design; *Nat. Med.* 4, 1366–1367.
25. Mergny, J.L., Lacroix, L., Teulade-Fichou, M.-P., Hounsou, C., Guittat, L., Hoarau, M., Arimondo, P.B., Vigneron, J.-P., Lehn, J.-M., Riou, J.-F., Garestier, T. and Helene, C. (2001) Telomerase inhibitors based on quadruplex ligands selected by a fluorescence assay; *Proc. Natl. Acad. Sci. USA* 98, 3062–3067.
26. Riou, J.-F., Guittat, L., Mailliet, P., Laoui, A., Renou, E., Petitgenet, O., Megnin-Chanet, F., Helene, C. and Mergny, J.L. (2002) Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands; *Proc. Natl. Acad. Sci. USA* 99, 2672–2677.
27. Arora, A., Kumar, N., Agarwal, T. and Maiti, S. (2010) Human telomeric G-quadruplex: targeting with small molecules; *FEBS J.* 277, 1345–1349.
28. Zhang, J.-L., Fu, Y., Zheng, L., Li, W., Li, H., Sun, Q., Xiao, Y. and Geng, F. (2009) Natural isoflavones regulate the quadruplex–duplex competition in human telomeric DNA; *Nucl. Acids Res.* 37, 2471–2482.
29. Jin, Y., Li, H. and Liua, P. (2010) Label-free electrochemical selection of G-quadruplex-binding ligands based on structure switching; *Biosens. Bioelectron.* 25, 2669–2674.
30. Jin, Y., Chen, G. and Wang, Y. (2011) Gold nanorod-based FRET assay for selection of G-quadruplex-binding ligands; *Gold Bull.* 44, 163–169.
31. Nakamura, Y., Yogosawa, S., Izutani, Y., Watanabe, H., Otsuji, E. and Sakai, T. (2009) A combination of indol-3-carbinol and genistein synergistically induces apoptosis in human colon cancer HT-29 cells by inhibiting Akt phosphorylation and progression of autophagy; *Mol. Cancer* 8, 100–115.
32. Li, Y., Upadhyay, S., Bhuiyan, M. and Sarkar, F.H. (1999) Induction of apoptosis in breast cancer cells MDA-MB-231 by genistein; *Oncogene* 18, 3166–3172.

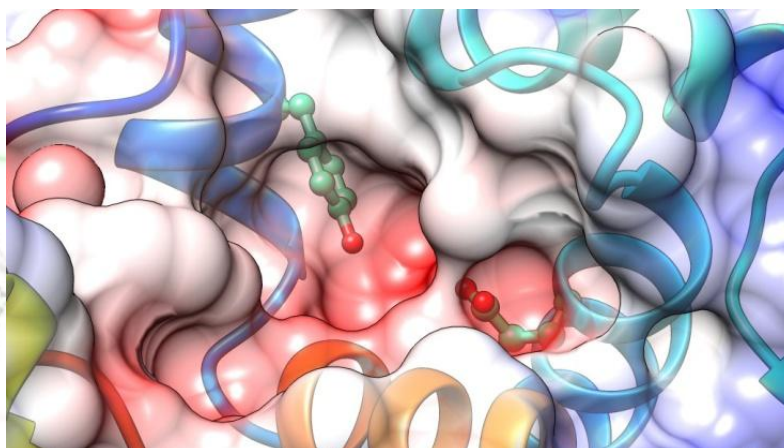
33. Sakamoto, T., Horiguchi, H., Oguma, E. and Kayama, F. (2009) Effects of diverse dietary phytoestrogens on cell growth, cell cycle and apoptosis in estrogen-receptor-positive breast cancer cells; *J. Nutr. Biochem.* 21, 856–864.
34. Sathyamoorthy, N. and Wang, T.T. (1997) Differential effects of dietary phyto-oestrogens daidzein and equol on human breast cancer MCF-7 cells; *Eur. J. Cancer* 33, 2384–2389.
35. Peterson, G. and Barnes, S. (1993) Genistein and Biochanin-A inhibit the growth of human prostate cancer cells but not epidermal growth factor receptor tyrosine autophosphorylation; *Prostate* 22, 335–345.
36. Kim, S.H., Kim, Y.B., Jeon, Y.T., Lee, S.C. and Song, Y.S. (2009) Genistein inhibits cell growth by modulating various mitogen-activated protein kinases and AKT in cervical cancer cells; *Ann. N. Y. Acad. Sci.* 117, 495–500.
37. Menon, L.G., Kuttan, R., Nair, M.G., Chang, Y.C. and Kuttan, G. (1998) Effect of isoflavones genistein and daidzein in the inhibition of lung metastasis in mice induced by B16F-10 melanoma cells; *Nutr. Cancer* 30, 74–77.
38. Das, A., Banik, N.L. and Ray, S.K. (2009) Flavonoids activated caspases for apoptosis in human glioblastoma T98G and U87MG Cells but not in human normal astrocytes; *Cancer* 116, 164–176.
39. Spinuzzi, F., Pagliacci, M.C., Migliorati, R., Grignani, M.F. and Riccardi, C. (1994) The natural tyrosine kinase inhibitor genistein produces cell cycle arrest and apoptosis in Jurkat T-leukemia cells; *Leuk. Res.* 18, 431–439.
40. Yeh, T.C., Chiang, P.C., Li, T.K., Hsu, J.L., Lin, C.J. and Wang, S.W. (2007) Genistein induces apoptosis in human hepatocellular carcinomas via interaction of endoplasmic reticulum stress and mitochondrial insult; *Biochem. Pharmacol.* 73, 782–792.
41. Moiseeva, E.P., Almeida, G.M., Jones, G.D. and Manson, M.M. (2007) Extended treatment with physiologic concentrations of dietary phytochemicals results in altered gene expression, reduced growth, and apoptosis of cancer cells; *Mol. Cancer Ther.* 6, 3071–3079.
42. Fotsis, T., Pepper, M., Adlercreutz, H., Hase, T., Montesano, R. and Schweigerer, L. (1995) Genistein a dietary ingested isoflavonoid inhibits cell proliferation and *in vitro* angiogenesis; *J. Nutr.* 125, 790–797.
43. Magee, P.J., Raschke, M., Steiner, C., Duffin, J.G., Pool-Zobel, B.L., Jokela, T., Wahala, K. and Rowland, I.R. (2006) Equol: A comparison of the effects of the racemic compound with that of the purified S-enantiomer on the growth, invasion and DNA integrity of breast and prostate cells *in vitro*; *Nutr. Cancer* 54, 232–242.
44. Jagadeesh, S., Kyo, S. and Banerjee, P.P. (2006) Genistein represses telomerase activity via both transcriptional and posttranslational mechanisms in human prostate cancer cells; *Cancer Res.* 66, 2107–2115.
45. Guo, J.M., Kang, G.Z., Xiao, B.X., Liu, D.H. and Zhang, S. (2004) Effect of daidzein on cell growth cell cycle and telomerase activity of human cervical cancer *in vitro*; *Int. J. Gynecol. Cancer* 14, 882–888.
46. Chemical Computing Group, Inc. Molecular Operating Environment Release 10. Montreal, Canada, 1998. <http://www.chemcomp.com> (Accessed July, 2011).
47. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. and Olson, A.J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility; *J. Comput. Chem.* 16, 2785–2791.
48. Lang, P.T., Brozell, S.R., Mukherjee, S., Pettersen, E.T., Meng, E.C., Thomas, V., Rizzo, R.C., Case, D.A., James, T.L. and Kuntz, I.D. (2009) Dock6: Combining technique to model RNA-small molecule complexes; *RNA* 15, 1219–1230.
49. Huiyong, S., Youyong, L., Mingyun, S., Sheng, T., Lei, X., Peichen, P., Yan, G. and Tingjun, H. (2014) Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring; *Phys. Chem. Chem. Phys.* 16, 22035–22045.

50. Mitrasinovic, P.M., Tomar, J.S., Nair, M.S. and Barthwal, R. (2011) Modeling of HIV-1 TAR RNA-ligand complexes; *Med. Chem.* 8, 301–308.
51. Holt, P.A., Chaires, J.B. and Trent, J.O. (2008) Molecular docking of intercalators and groove-binders to nucleic acids using AutoDock andSurflex; *J. Chem. Inf. Model.* 48, 1602–1615.
52. Ricci, C.G. and Netz, P.A. (2009) Docking studies on DNA-ligand interactions: Building and application of a protocol to identify the binding model; *J. Chem. Inf. Model.* 49, 1925–1935.
53. Hou, J.-Q., Chen, S.-B., Zan, Li.-P., Ou, T.-M., Tan, J.-H., Leonard, G.L. and Huang, Z.-S. (2015) Identification of a selective G-quadruplex DNA binder using a multistep virtual screening approach; *Chem. Commun.* 51, 198–201.
54. Lu, X.-J. and Wilma, O.K. (2003) X3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures; *Nucl. Acids Res.* 31, 5108–5121.
55. Jain, A.K. and Bhattacharya, S. (2011) Interaction of G-quadruplexes with non-intercalating duplex-DNA minor groove binding ligands; *Bioconjugate Chem.* 22, 2355–2368.
56. Wang, B.H., Ternai, B. and Polya, G. (1997) Specific inhibition of cyclic AMP-dependent protein kinase by warangalone and robustic acid; *Phytochemistry* 44, 788–796.
57. Sasaki, K., Tsurumaru, Y., Yamamoto and H. Yazaki, K. (2011) Molecular characterization of a membrane-bound prenyltransferase specific for isoflavone from *sophoraflavescens*; *Biol. Chem.* 286, 24125–24134.
58. di Leva, F.S., Novellino, E., Cavalli, A., Parrinello, M. and Limongelli, V. (2014) Mechanistic insight into ligand binding to G-quadruplex DNA; *Nucl. Acids Res.* 42, 5447–5455.
59. Ou, T.-M., Lu, Y.-J., Tan, J.-H., Huang, Z.-S., Wong, K.-Y., and Gu, L.-Q. (2008) G-Quadruplexes: Targets in anticancer drug design; *Chem. Med. Chem.* 3, 690–713.
60. Harrison, R.J., Reszka, A.P., Haider, S.M., Romagnoli, B., Morrell, J., Read, M.A., Gowan, S.M., Incles, C.M., Kelland, L.R. and Neidle, S. (2004) Evaluation of by disubstitutedacridone derivatives as telomerase inhibitors: the importance of G-quadruplex binding; *Bioorg. Med. Chem. Lett.* 14, 5845–5849.
61. Sun, Y.R., Elzbieta, I., Richard, L., Karen, D., Daekyu, S., Mary, P.M., Roodman, G.D., Lawrence, H. and Daniel, V.H. (2000) Effect of Telomere and Telomerase Interactive Agents on Human Tumor and Normal Cell Lines; *Clin. Cancer Res.* 6, 987–993.
62. Parkinson, G.N., Ghosh, R., and Neidle, S. (2007) Structural Basis for Binding of Porphyrin to Human Telomeres; *Biochemistry* 46, 2390–2397.

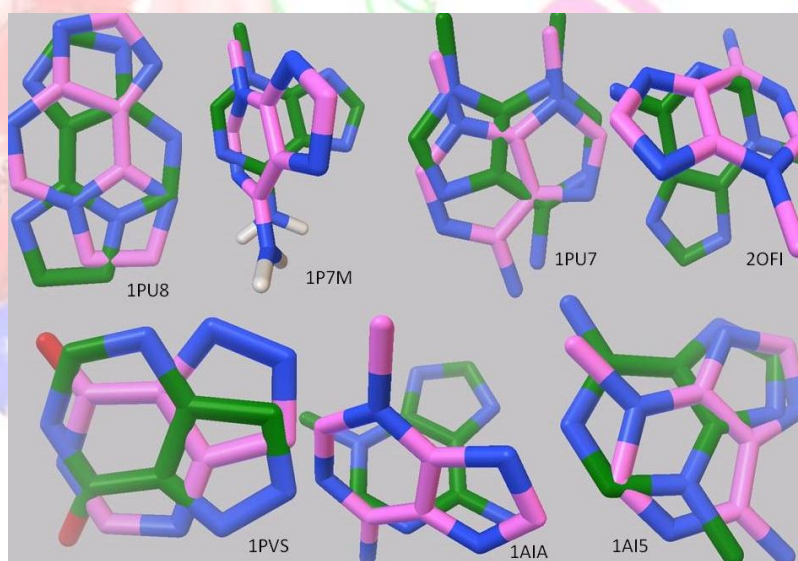
APPENDICES



Appendix I-1. Genetic aspects of acquisition of antibiotic resistance (Bacteria can become antibiotic resistant (Ab^r) by mutation of the target gene in the chromosome). They can acquire foreign genetic material by incorporating free DNA segments into their chromosome (transformation). Genes are also transferred following infection by bacteriophage (transduction) and through plasmids and conjugative transposons during conjugation. The general term transposable element has been used to designate (1) an insertion sequence, (2) composite (compound), complex, and conjugative transposon, (3) transposing bacteriophage, or (4) integron.



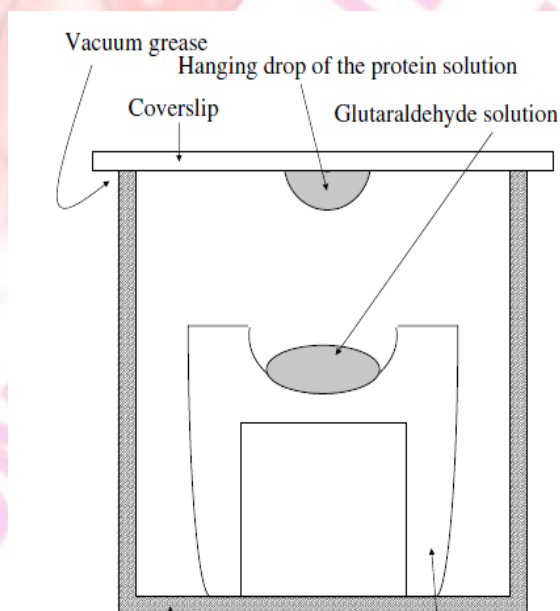
Appendix II-3. Active site of model showing electrostatic surface. Red colour of pocket indicates the electron richness favourable for the electron-deficient 3mA. H-bond forming residues Tyr and Glu are shown as ball and stick.



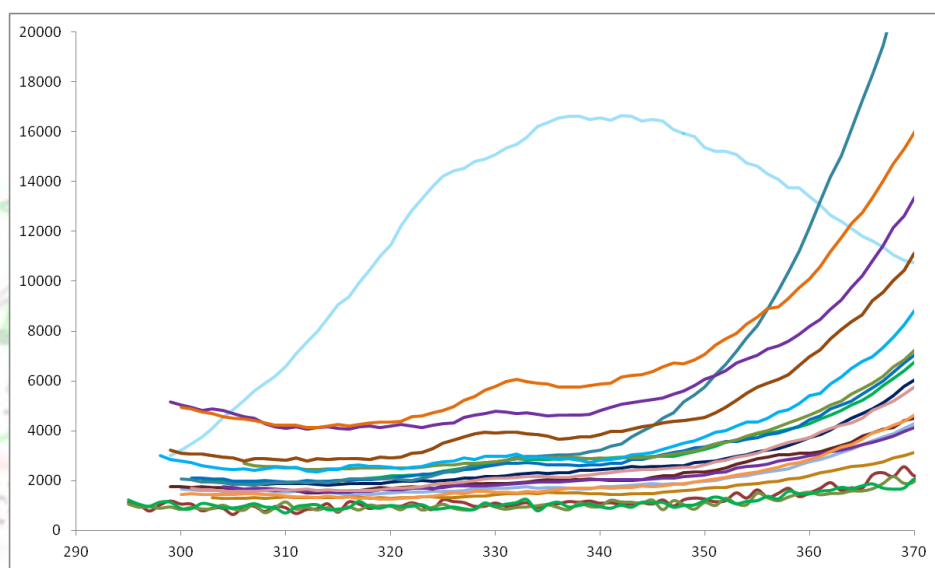
Appendix II-4. The output docking poses of ligands (PDB id. 2OFI, 1PVS, 1PU7, 1PU8, 1P7M, 1AIA and 1AI5) superimposed on the experimental pose (Green colour: Experimental, Pink colour: Theoretical).

Chains	Interface residues (no.)	Interface area (Å ²)	H-bonds (no.)	non-bonded interactions (no.)	ΔG, kcal/mol	Buried area, Å ²
A=C	07:07	490:480	2	043		
A=D	21:21	935:933	6	118	-19.5	1860
A=G	04:05	211:183	0	011		
B=C	20:21	929:930	5	100	-19.5	1860
B=D	07:08	370:358	4	044		
D=H	10:11	525:545	5	098		
E=G	08:07	470:480	0	051		
E=F	20:22	933:941	6	105	-19.0	1870
F=H	03:03	138:136	1	004		
G=H	22:19	956:960	5	108	-19.2	1910

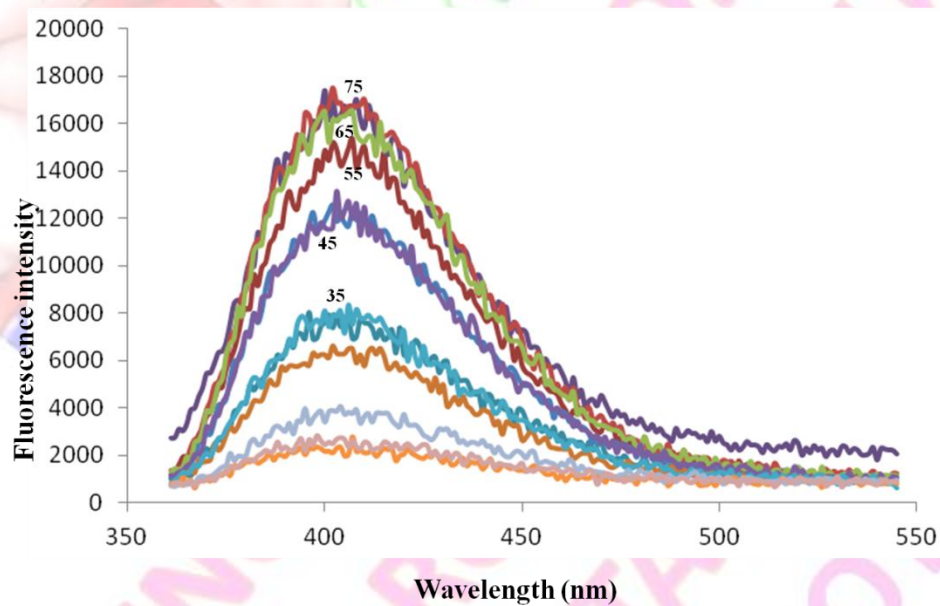
Appendix III-1. The interaction statistics of different chains of the pdb-1xeb. The ΔG and buried surface area are calculated using PISA server for the chains which exist as stable dimer in the solution.



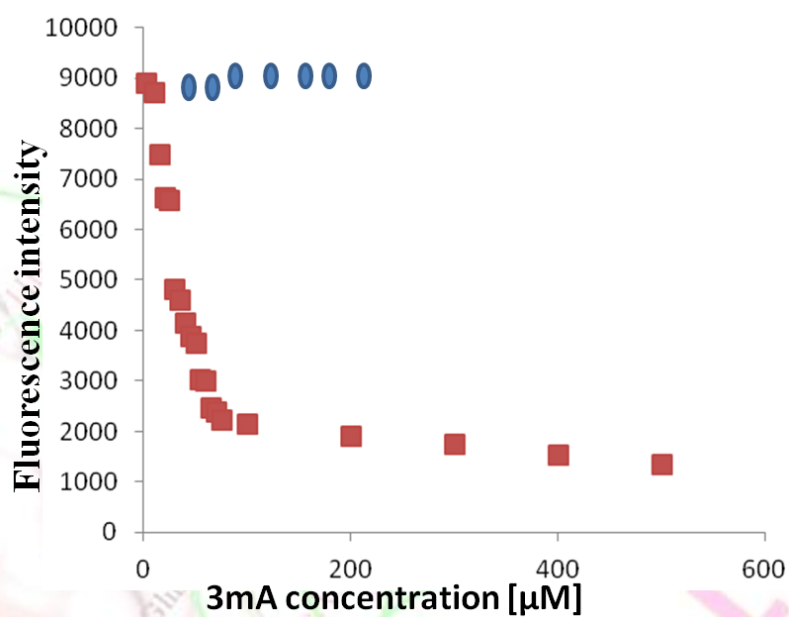
Appendix III-2. Schematic representation of experimental setup used in the glutaraldehyde crosslinking assay.



Appendix IV-1. Titration of TAG (2 μM) with the 3mA (2–500 μM).



Appendix IV-2. Extrinsic thermal denaturation of TAG.



Appendix IV-3. Tryptophan fluorescence studies TAG binding to 3mA (Red square marker) and adenine (Blue round marker).

Incubation time (min)	Peak area (3mA)	Peak intensity (3mA)	% of Peak area (3mA)
0	0	0	0
10	207876	2496	7.23
20	554443	3751	20.04
30	5458562	24146	75.59
40	6691799	34819	76.38
50	5720154	31019	75.91
60	3615485	18250	73.16
70	2882102	60885	80.57
80	5202681	37638	81.06
90	10526328	66970	90.04
100	10870199	70827	90.15
110	10871485	67010	90.35
120	10872156	68581	90.48

Appendix IV-4. Glycosylase assay parameters obtained using HPLC experiment.

LIST OF PUBLICATIONS

Journals:

1. Petar M. Mitrasinovic, Jyoti Singh Tomar, Maya S. Nair, Ritu Barthwal
Modeling of HIV-1 TAR RNA-Ligand Complexes.
Medicinal Chemistry, 2011 (7) 301–308.
2. Jyoti Singh Tomar
In silico modeling studies of G-quadruplex with soy isoflavones having anticancerous activity.
Journal of Molecular Modeling, 2015, DOI: 10.1007/s00894-015-2723-0.
3. Jyoti Singh Tomar, R. K. Peddinti
Optimized method for TAG protein homology modeling: *In silico* and experimental structural characterization.
International Journal of Biological Macromolecules, 2016 (*accepted*).
4. Jyoti Singh Tomar, R. K. Peddinti
Modified homology modelling: Construction and analysis of protein–protein interactions in *A. baumannii* histone acetyl transferase Hpa2.
Journal of Biomolecular Structure and Dynamics, 2016 (*accepted*).
5. Jyoti Singh Tomar, P. Kumar, R. K. Peddinti
DNA binding TAG protein from *A. baumannii*: Biophysical and biochemical characterization.
(*Manuscript in preparation*).
6. Jyoti Singh Tomar, R. K. Peddinti
Expression, purification and biophysical characterization of the protein histone acetyl transferase Hpa2.
(*Manuscript in preparation*).

Conferences:

1. “Binding of berberine to poly (dA-dT) and the promoter sequence d(CCAATTGG)₂: experimental and computational evidence.” In Meeting of the Indian Biophysical Society (New Delhi), January 30–February 2, 2011.
2. “Interaction of protoberberine alkaloid with promoter sequence d(CCAATTGG)₂.” In Meeting of the Indian Biophysical Society (New Delhi), January 30–February 2, 2011.
3. Workshop on “Recent advances in NMR spectroscopy” from December 12–17, 2011 in the TIFR Hyderabad.
4. “Modeling of HIV-1 TAR RNA-ligand complexes using Dock 6 .4” In CoFIBS, October 3–5, 2010 at National Institute of Technology, Rourkela, Orissa, India.
5. “Docking study of minimized structure of Anatoxin-A: A potent neurotoxin, ABA 2012.
6. “Optimized efficient method for TAG protein structural and functional characterization using *in silico* and experimental techniques.” In 22nd conference of national magnetic resonance society, India. February 18–21, 2016 at Department of Chemistry, IIT Kharagpur.