The heart is a truly remarkable organ, beating more than 40 million times a year and pumping over 7500 liters of blood a day. The disease caused to the heart and blood vessel or vascular system is called cardiovascular disease (CVD) [Robbins et al., 2012].

CVD is thought to be a disease of modern human beings and related to present lifestyles. But Thompson et al., [2013] identified hardening of the arteries across 4000 years of human history which is today known as Atherosclerosis. In their report they performed the whole body Computerized tomography (CT) scans of 137 mummies from four different geographical regions or populations, and interpreted that Atherosclerosis was common in four populations which were pre-industrial and pre-agricultural hunter gatherers. CVDs are the number one cause of death globally, more people die annually from CVDs than from any other cause. CVDs succeeded across 4000 years earlier [Thompson et al., 2013] and even today they are the number one cause of death globally [WHO, 2011].

Blood is normally a sterile environment. Hence, the detection of bacteria in the blood is always abnormal and is called Bacteremia. [Ochei and Kolhatkar, 2000]. Silent bacteraemia is caused by the non-pyogenic group of streptococci which are reported for causing pus-less infections with the human host (Health Protection Agency, UK).

2.1 Infective Cardiovascular disease - Rheumatic heart disease and infective endocarditis

Earlier it was presumed that the CVD is one of the non-infective disease causing severe ailments related to heart. Rheumatic Heart Disease
(RHD) is a kind of CVD and a form of inflammation of the inner tissue of the heart caused by infectious agents [Bonow et al., 2012]. The illness typically develops two to three weeks after an infection. The illness is so named because of its similarity in presentation to rheumatism [Dorland, 2011]. RHD is the beginning of infective endocarditis (IE). Normally, blood flows smoothly through heart valves. If they have been damaged from rheumatic fever, the risk of bacterial attachment is increased [Kumar et al., 2007].

In developing countries like India where tradition is more important than western guidelines, recurrence of the rheumatic heart fever/rheumatic heart disease (RHF/RHD) is most common predisposing factor indicating IE [Sadiq et al., 2001; Senthilkumar et al., 2010]. On contrary, reports from the West have shown a change in the spectrum of IE [Ferreiros et al., 2006; Tornos et al., 2005]. However, reports from India are very few, and those too are probably not a true reflection of the spectrum of IE due to a lack of the diagnostic and management modalities in the country [Gupta et al., 2013].

2.2 Infective Endocarditis

Infective Endocarditis (IE) is a (usually bacterial) infection of the endocardium or equivalent prosthetic surfaces in the heart. The incidence of IE and its mortality has not changed significantly over many years [Kiefer and Bashore, 2012; Murdoch et al., 2009]. Without treatment, IE is usually lethal. Even with surgical treatment (used in about 50% of patients) and antibiotics, the mean in-hospital mortality is as high as 15–20%, and approximate mortality of 1 year is up to 40% [Murdoch et al., 2009]. Even with available antibiotic therapies, mortality due to IE remains at levels approaching 25% [Sy et al., 2011].

Durack et al., [1994] designed a study to develop improved criteria for the diagnosis of infective endocarditis and to compare these criteria
with then accepted criteria in a large series of cases. This was named as ‘Duke criteria’. They concluded by claiming that the application of this criteria would increase the number of definite diagnoses of IE. The Duke criteria are a collection of major and minor criteria used to establish a diagnosis of infective endocarditis. Later, the Duke criteria was revised for more improved diagnosis [Li et al., 2000].

An annoying fact is the increase in drug resistance by several etiological agents. *S. aureus* has grown to be a methicillin resistant and also the emergence of daptomycin resistance in *E. hirae*, *Enterococci*, and several other oral streptococci [Que and Moreillon, 2011; Opal and Pop-Vicas, 2010], whose clinical incidences for patients with IE has been well reported.

**2.3 Pathogenesis**

IE typically affects the heart valves. However, it can also occur in the interventricular septum, the mural endocardium, the chordae tendineae, and other structures of heart [Freedman, 1982; Smilack, 1983]. The most common sites of endocarditis are left sided, typically involving the mitral valve. Weakening of the endocardium, a major risk factor for the pathogenesis of IE, is thought to have three major causes [Kumar et al., 2007; Wilson et al., 2007] as follows:

1. Uneven blood flow – caused by valvular incompetence
2. Direct physical damage – caused by trauma
3. Degeneration – brought on by age or other disease processes

Pathogens gain a provisional access to the bloodstream as a result of dentogen pathway, health-care procedures, or by intravenous drug use (fig.-RL1). Once these pathogens enters the bloodstream they can rapidly (within minutes) adhere via platelet fibrin deposition to a mechanically injured valve surface due to pre-existing valvular disease or to a valve surface with inflammation. Later, such adherence within the host leads to
proliferation and maturation of the colonies on the valve [Werdan et al., 2014]. Consequently, such colonization and haematogenous universal spreading of the pathogens often occurs, leading to complications such as ischaemic stroke, cerebral haemorrhage, meningitis or meningeal reaction, brain abscess, and mycotic aneurysm [Fowler et al., 2010]. Some bacterial pathogens, including S. aureus, have even been recognized as facultative intracellular pathogens [Sendi et al., 2009; Fraunholz and Sinha, 2012].

**Mechanism**

Werdan et al., [2014] have reviewed the mechanism of establishment of pathogens IE. For bacterias to get colonized in host, circulating pathogens in the blood must escape the bloodstream and settle to establish an infection. There are two mechanisms for colonization to occur. A consequence of either mechanical injury or endothelial inflammation, provides an opportunity for pathogens to colonize (fig.-RL2).

Valve colonization as a consequence of mechanical injury occurs when bacteria bind to coagulum and colonize it during temporary bacteraemia (fig.-RL2 (A-1)). Later, adhered monocytes release tissue factor and cytokines (fig.-RL2 (A-2)). Nest to this more platelets are attracted and become activated, and the bacterial vegetation grows (fig.-RL2 (A-3)). Further, endothelial cells are infected, and can be lysed by bacterial products, or bacteria can persist inside the cells (fig.-RL2 (A-4)).
Valve colonization as a consequence of an inflammatory endothelial lesion occurs when activated endothelial cells express integrins that promote the local deposition of fibronectin; bacteria such as *Staphylococcus aureus* adhere to this protein (fig.-RL2 (B-1)). Bacteria are internalized, and endothelial cells release tissue factor and cytokines, causing blood clotting and promoting the extension of inflammation and vegetation formation (fig.-RL2 (B-2)). Infected endothelial cells can be lysed by bacterial products, or bacteria can persist inside the cells by changing the phenotype like developing drug resistivity (fig.-RL2 (B-3)).

Fig.-RL2: Mechanisms of infective endocarditis [Courtesy: Werdan et al., 2014].
2.3.1 Infective endocarditis and viridans group streptococci

IE has a large number of etiological agents and streptococci account for 50%-80% of IE cases [Murdoch et al., 2009]. The viridans group streptococci (VGS) are a heterogeneous group of organisms that can be human commensals, colonizing the gastrointestinal and genitourinary tracts in addition to the oral mucosa [DeSimone et al., 2012]. Silent bacteraemia by VGS occur via dentogenu pathway especially in individuals with poor dental hygiene. This can result in endocarditis amongst vulnerable individuals, especially those with underlying conditions such as malignancies, diabetes or chronic heart disease and also with increasing age [Elshibly et al., 2014; Souza et al., 2014; Weening-Verbree et al., 2013]. VGS are the major causes for streptococcal infective endocarditis [Kumar et al., 2013].

Non-pyogenic and VGS group of streptococci are inhabitants of the oral cavity, which is sterile at birth, but at the age of one year, these species compose 70% of the cultivable oral flora [McCarthy et al., 1965; Smith et al., 1993]. Severe infections caused by them lead to IE in addition to odontofacial infections, brain abscesses and abdominal infections [Westling, 2005].

2.4 Viridans group streptococci and taxonomy

Few microorganisms turn to versatile pathogens. Genome inconsistency in bacteria reflects evolutionary, developmental and adaptive traits leading resistance towards new drugs and environments [Larsson et al., 2011]. Historically, VGS have been troublesome organisms to accurately characterize and identify [Doern and Burnham, 2010]. Identification of VGS to the species level is difficult because VGS exchange genetic material [Simmon et al., 2008]. Struggle has been made
in various fields of science to automate and develop the classification system of such bacterial groups. For example, techniques ranging from genome-based classification to bacterial culture images and Bayesian semi-supervised model based tools have been available for species level identification and classification of bacterial pathogens [Klenk and Göker 2010; Dubuisson et al., 1994; Cheng et al., 2011].

Carl Woese and George E. Fox were two of the scientists who pioneered the use of 16S rRNA in phylogenies. They reasoned the use of 16S rRNA for reconstructing phylogenies as due to slow evolution rates of genes (referred as 16S rDNA) coding for 16S rRNA [Woese and Fox, 1977]. Using the 16S rRNA phylogenies Kilian et al., [2005], classified the streptococci into six groups. They were grouped as pyogenic group, mitis group, mutans group, anginosus group, salivarius group and bovis group. Adding an exception to his own ideas of phylogenies, Kilian et al., [2009] found that one of the leading pathogens affecting mankind, S. pneumoniae and even a member of Mitis group, is strikingly similar to the three commensal species S. mitis, S. oralis, and S. infantis. Such similarities often cause problems of identification in clinical microbiology laboratories [Whatmore et al., 2000; Hanage et al., 2005; Carvalho et al., 2003]. Further, he emphasized on this problem saying this was due to introduction of the species S. pseudopneumoniae into mitis [Arbique et al., 2004]. The first phylogenetic relationships of VGS specifically causing endocarditis was first performed by Simmon et al., [2008]. They isolated VGS specifically from IE patients and used tuf, and rpoB gene targets and developed the 16S rRNA phylogeny.

Since the more-conserved rRNA gene sequences do not always allow species discrimination, genome based multi-gene sequence comparison is one of the promising tools to analyze bacterial populations [Sawabe et al., 2013; Martens et al., 2008]. In contrast to native 16S rRNA gene sequence analysis, methods like multilocus sequence analysis (MLSA) is capable of yielding sequence clusters at a wide range of
taxonomic levels, from species level to clusters at higher levels [Gevers et al., 2005]. MLSA even offers considerable advantages of reproducibility of procedure and allows the generation of cumulative databases [Maiden et al., 1998; Rong et al., 2010]. One of such classification have also been adapted for VGS group and even have made use of the internet to assign the bacterial strains to their respective species whose species boundaries are not well known [Bishop et al., 2009].

Thompson et al., [2013] explained the streptococcal taxonomy based on complete genome sequence analyses. They later defined a 
Streptococcus
species as a group of strains that share ≥ 95% DNA similarity in MLSA and average amino acid identity, and > 70% DNA identity in genome-to-genome distances.

2.5 Comparative genomics and viridans group streptococci

Comparative genome analysis among organisms of the same species not only shows a high degree of similarity in gene content and organization, but also a gives an idea of sequence heterogeneity as evidenced by the existence of large number of single nucleotide polymorphisms for their persistence and virulence [Krupovic et al., 2011]. Many reports have been published on the aspects of comparative genomics of the genus 
Streptococcus
to understand the biology of evolution. For example, Ferretti et al. [2004] by comparative genome analysis between the several streptococcal species showed that 

S. pyogenes
was more closely related to 

S. agalactiae
than with 

S. pneumoniae
or 

S. mutans.

Evan many reports are related to comparative genome analysis within the Mitis group to understand the foremost purpose of their virulence by either commensalism, sharing of some specific genes, antagonism, etc. all of which are for survival and growth. Hiller et al., [2007] have sequenced the genomes of 17 

S. pneumoniae
strains, belonging
to Mitis group. They ultimately found 46% of core set of genes and 54% of dispensable genes. The comparative genome analysis of a Mitis strain *S. parasanguinis* FW213 indicated that its genome is shaped by chromosomal inversion, recombination and horizontal gene transfer (HGT) events. This also explained the basic information on the physiology and potential pathogenic capacity of the bacterium to maintain an ecological niche in dental plaque, escape from host defense and establish infection in heart valves. [Geng et al., 2012]. Similarly, Shahinas et al., [2013] have used comparative genomics approaches to understand the virulence and dynamics of commensalism among VGS. They have identified similarities and key differences between *S. pneumonia* and *S. mitis* with that of *S. pseudopneumoniae* IS7493. They have observed that the genome structure of *S. pseudopneumoniae* IS7493 is most closely related to that of *S. pneumoniae* R6, but with the evidence of several recombination. Additionally, they also have found that *S. pseudopneumoniae* has several virulence factors and shows antibiotic resistance mechanisms hence suggesting the sharing of such genes increases the antibiotic resistance levels seen among the VGS. Comparative genomics approach was also applied to yield a better understanding of the evolution of the genus *Streptococcus* by Gao et al., [2014]. They considered several factors such as genome dynamics, population structure, phylogenies and virulence factor distribution among species groups. This study among VGS, also indicated that the pan-genome size increases with the addition of newly sequenced strains, while the core genome size decreases with sequential addition at the genus level and species group level, hence conforming the hypothesis as proposed by Tettelin et al., [2005].

The concept of synteny has been applied for finding structural variations between closely related genomes [Kaper et al., 2004]. Finding syntenies shared among many strains within the same microbial species helps to infer the minimum genomic material (or core genome) required for bacterial life and thus can be useful for identifying minimal gene set
[Gibson et al., 2008]. Ghiurcuta and Moret [2014] have evaluated the use of syntenic comparison in comparative genomics and have identified its use in comparison of close species. They have also proposed that syntenic blocks facilitates comparative genomic studies in terms of robustness and accuracy.

Currently, comparative genomics has led path to a new concept of subtractive genomics. These methods are being widely used to identify species specific novel drug and vaccine targets. If the researcher is unable to advance with experimental methods because of some technical constraints or with an aim to reducing costs and increasing research efficiency, one can use such theoretical approaches to cater the need of identifying novel targets by in silico methods. Such a subtractive genomics study was implemented very recently by Chowdhury et al., [2014] to reveal the novel drug targets of S. sanguinis, where they have identified 15 proteins that are unique in several metabolic pathways of S. sanguinis.

The fundamental principle of the subtractive genomic approach is “a good therapeutic target is a gene/protein essential for the pathogen survival, but which cannot be found in host” [Barh et al., 2011]. Along with this the successful completion of the largest collaborative biological project that is human genome project [Venter et al., 2001; Lander et al., 2001] has helped to identify novel drug and vaccine targets in human pathogens.

Since there is a need for identification of such novel targets within drug resistant pathogens and which are not amiable for the purpose, theoretical approaches cater the need of a researcher through ‘in silico’ methods [Pucci et al., 2006].

Both S. gordonii and S. sanguinis are primary colonizers [Xu et al., 2014] as represented in fig.-RL3. The genome sequence of S. gordonii strain
Challis CH1 is deposited in GenBank [Vickerman et al., 2007]. Also the genome of *S. sanguinis* SK36 is available [Xu et al., 2007]. This facilitated the comparison of genome of *S. gordonii* marching towards the identification of unique druggable targets.

Fig.-RL3: The tooth surface covered by various oral bacterial colonizers, representing *S. gordonii* as one of the primary colonizers [Courtesy: Bakaletz et al., 2004].

### 2.5.1 Unique essential proteins and Domain of unknown functions

Lateral gene transfer is the driving force to acquire new genes by a bacteria, which are unique and such acquirement is due to the gene essentiality [Doolittle, 1999]. Xu et al., [2011] have presented a clear perception of essential genes in *S. sanguinis* which is a prominent causative isolate of IE. They identified 218 essential genes using single-
gene knockout method. Their careful examination revealed that these essential genes were associated with only three basic categories of biological functions: maintenance of the cell envelope, energy production, and processing of genetic information. Later, also the essential genes of *S. pneumoniae* were identified which were potentially required for growth and survival in human saliva. The study lead to the identification of 147 genes performed by using transposon sequencing method. Among these essential genes, they were predicted to be involved in cell envelope biosynthesis, cell transport, amino acid metabolism, and stress response predominated [Verhagen et al., 2014].

Unique domains in essential proteins can be found in bacteria, whose functions are unidentified and are called as Domain of unknown functions (DUFs). Such proteins which are essential but whose functions are unidentified are known as eDUFs. About 2,700 DUFs are found in bacteria compared with just over 1,500 in eukaryotes. Over 800 DUFs are shared between bacteria and eukaryotes, and about 300 of these are also present in archaea. It is found that more than 20% of all protein domains were annotated as DUFs which constitutes eDUFs of bacteria [Goodacre et al., 2014].

Many factors contribute to the virulence of a pathogen. Pallen et al., [2007] stated that “It has also become evident that even the definitions for 'pathogen' and 'virulence factor' need to be re-evaluated”.

### 2.6 Protein-protein interaction

In the 1950s, regarding the structural aspects of protein, there was no absolute information regarding its conformation. It is fortunate for biologists that this all changed dramatically, reasonably and rapidly, due to invent of structural biology [Banaszak, 2000]. The first attempt to describe the characteristics of protein interaction sites was undertaken in 1975 by Chothia and Janin. With data from only three complexes, they suggested that the residues which form the interface are closely packed,
tend to be hydrophobic and that complementarity may be an important factor in predicting which proteins can interact. Now a days it is clear that protein-protein interactions have a key role in most biological processes, and offer attractive opportunities for therapeutic intervention [Zinzalla and Thurston, 2009; Milroy et al., 2014]. Developing small molecules that modulate protein-protein interactions is difficult, but possible. After genome reduction, it remains with unique protein sequences that can be potent and unique drug target [Arkin and Wells, 2004]. It is very significant to find out the most potential metabolic functional association in any pathogen to predict a potent drug target [Snel et al., 2000].

Experimental methods, including high-throughput techniques are highly resource intensive. Therefore, computational discovery of PPIs can accelerate biological discovery by presenting most promising pairs of proteins that are likely to interact. For many bacteria, genome sequence, and thereby genomic context of proteomes, is readily available; additionally, for some of these proteomes, localization and functional annotations are also available, but interactomes are not available. PPI network studies are contributing to find precisely the important proteins/enzymes, which interact and play a role of pathogenicity in many infectious and systemic diseases which can be concluded as potential drug targets [Kushwaha and Shakya, 2010].

Search Tool for the Retrieval of Interacting Genes (STRING) is a database of such PPI prediction. The interactions include direct (physical) and indirect (functional) associations, it provides uniquely comprehensive coverage which includes interactions from published literature describing experimentally studied interactions along with genome analysis using several established methods such as domain fusion, phylogenetic profiling and gene neighborhood concepts. It provides ease of access to both experimental as well as predicted interaction information [Szklarczyk et al., 2011]. Analysis of such
interactions provided both verification of known virulence factors and identification of new putative virulence proteins [Jensen et al., 2009].

Proper analysis and identification of hub protein among the set of unique proteins gives an idea to perform simulation studies like protein modeling and protein ligand docking. Targeting protein–protein interfaces of regulatory multiprotein complexes has become a significant focus in drug discovery [Jubb et al., 2012].

2.7 Phosphoenolpyruvate-dependent Phosphotransferase system pathway

Phosphoenolpyruvate-dependent Phosphotransferase system (PEP) group translocation, also known as the phosphotransferase system (PTS), is a distinct method used by bacteria for sugar uptake where the source of energy is from phosphoenolpyruvate (PEP). It is known as multicomponent system that always involves enzymes of the plasma membrane and those in the cytoplasm. An example of this transport is found in E. coli cells. The system was discovered by Saul Roseman in 1964. Many researchers have explained the pathway of PEP:PTS in both gram positive and gram negative bacteria.

Numerous gram-positive bacteria take up carbohydrates through the PEP:PTS (fig.-RL4). This system transports and phosphorylates carbohydrates at the expense of PEP and is the subject of this review. The PTS consists of two general proteins, enzyme I and HPr, and a number of carbohydrate-specific enzymes, the enzymes II. PTS proteins are phosphoproteins in which the phospho group is attached to either a histidine residue or, in a number of cases, a cysteine residue. After phosphorylation of enzyme I by PEP, the phospho group is transferred to HPr. The enzymes II are required for the transport of the carbohydrates across the membrane and the transfer of the phospho group from phospho-HPr to the carbohydrates. Biochemical, structural, and molecular genetic studies have shown that the various enzymes II have
the same basic structure. Each enzyme II consists of domains for specific functions, e.g., binding of the carbohydrate or phosphorylation. Each enzyme II complex can consist of one to four different polypeptides. The enzymes II can be placed into at least four classes on the basis of sequence similarity. The genetics of the PTS is complex, and the expression of PTS proteins is intricately regulated because of the central roles of these proteins in nutrient acquisition. In addition to classical induction-repression mechanisms involving repressor and activator proteins, other types of regulation, such as antitermination, have been observed in some PTSs. Apart from their role in carbohydrate transport, PTS proteins are involved in chemotaxis toward PTS carbohydrates. Furthermore, the IIAGlC protein, part of the glucose-specific PTS, is a central regulatory protein which in its non-phosphorylated form can bind to and inhibit several non-PTS uptake systems and thus prevent entry of inducers. In its phosphorylated form, P-IIAGlC is involved in the activation of adenylate cyclase and thus in the regulation of gene expression. By sensing the presence of PTS carbohydrates in the medium and adjusting the phosphorylation state of IIAGlC, cells can adapt quickly to changing conditions in the environment. In gram-positive bacteria, it has been demonstrated that HPr can be phosphorylated by ATP on a serine residue and this modification may perform a regulatory function.

Proteins downstream of HPr tend to vary between the different sugars. The transfer of a phosphate group to the substrate once it has been imported through the membrane transporter prevents the transporter from recognizing the substrate again, thus maintaining a concentration gradient that favours further import of the substrate through the transporter.

With the glucose phosphotransferase system, the phosphorylation status of EIIA can have regulatory functions. For example, at low glucose concentrations phosphorylated EIIA accumulates and this activates membrane-bound adenylate cyclase. Intracellular cyclic AMP levels rise and this then activates CAP (catabolite activator protein), which is involved
in the catabolite repression system, also known as glucose effect. When the glucose concentration is high, EIIA is mostly dephosphorylated and this allows it to inhibit adenylate cyclase, glycerol kinase, lactose permease, and maltose permease. Thus, as well as the PEP group translocation system being an efficient way to import substrates into the bacterium, it also links this transport to regulation of other relevant proteins. It is an active transport. After the translocation, the metabolites transported are modified [Siebold et al., 2001; Postma et al., 1993].

![Diagram](image.png)

Fig.-RI.4: PEP:PTS in gram positive bacteria.

### 2.8 Protein druggability

Over the past 15 years, predictions over the ability of a protein to be a drug target have been a field of active research, and a variety of methods have been developed. Earlier to this, the methods relied on experimental druggability assessment [Hajduk et al., 2005; Cheng et al., 2007]. An example of such assessment includes NMR-based fragment screening performed by Hajduk and co-workers [2005]. The term druggability was first coined by Hopkins and Groom [2002], which describes a biological target (protein) that is known to or is predicted to bind with high affinity to a drug. The term gained its popularity after the completion of several genome projects with high throughput sequencing technologies [Hugo et al., 2012]. With the accessibility of complete human genome [Venter et al., 2001; Lander et al., 2001] and increasing availability pathogen genome
sequences, there always remains a challenge for substantial progress in structure determination of druggable proteins. Hence, the field of 'structural genomics' has emerged recently. Its aim is to characterize the structure of the genome [Griffiths et al., 2000]. Only a small number of proteins from pathogen have been structurally characterized to date. The only answer for the structures of remaining and experimentally undeterminable pathogen proteins can be provided by structural genomics. Using such methods, high-quality data from model organisms can be translated to under-studied pathogens based on orthology mapping [Agüero et al., 2008]. Later protein can be modelled based upon sequence homology and reliable models can be further carried with drug discovery process [Hillisch et al., 2004].

One recent review revealed that over 70 high-throughput biochemical screens against genetically validated drug targets in bacteria failed to yield even a single candidate that could be tested in the clinic [Payne et al., 2007]. The reasons for the failure of high throughput biochemical screens are not completely clear, but it could reflect the limited diversity of chemical libraries used or the absence of structural information for many of the targets [Van Voorhis et al., 2009]. These earlier studies generally show, the druggability of a given target in a pathogen can be predicted on the basis of factors such as the physicochemical nature of small-molecule binding sites on the target. The ease of developing an assay to screen for compounds that modulate the activity of the target protein is also an important consideration [Agüero et al., 2008].

2.8.1 Structure-based druggability

Early work on introducing some of the parameters of structure-based druggability came from Abagyan and coworkers [An et al., 2004] and then Fesik and coworkers [Hajduk et al., 2005], the latter by assessing the correlation of certain physicochemical parameters with hits from an NMR-based fragment screen. There has since been a number of publications
reporting related methodologies [Halgren, 2009; Schmidtke and Barril, 2010; Gupta et al., 2009]. Structure-based druggability is normally used to identify suitable binding pocket for a small molecule; however, some studies have assessed 3D structures for the availability of grooves suitable for binding helical mimetics [Jochim and Arora, 2010]. This is an increasingly popular approach in addressing the druggability of protein-protein interactions [Kozakov et al., 2011]. There exists several commercial tools and databases for structure-based druggability assessment. A publicly available database of pre-calculated druggability assessments for all structural domains within the Protein Data Bank (PDB) is provided through the ChEMBL’s DrugEBillity portal (https://www.ebi.ac.uk/chembl/drugability/).

Utilizing druggability assessments at large scale for aiding drug discovery is exemplified in the TDR Targets database [Agüero et al., 2008], where entire parasitic genomes were assessed for their druggability and biological essentiality to the pathogen in order to aid tropical disease drug discovery. Such druggability assessment was used by Cavalli and Bolognesi [2009] to identify novel lead candidates against Trypanosoma and Leishmania which cause neglected tropical diseases.

2.9 Virtual screening

Virtual screening of small-molecule libraries has been a rewarding computational approach that uses structural information of protein and molecular docking algorithms to facilitate drug discovery [Anderson, 2003; Guida, 1994; Tomlinson et al., 2009; Verlinde and Hol, 1994]. Molecular docking tools use algorithms which can often predict the conformation of a ligand bound to a protein’s active site with high accuracy [Trott and Olson, 2010; Kitchen et al., 2004; Lengauer and Rarey, 1996]. Virtual screening programs systematically perform molecular docking calculations on a data set (or library) of small molecules to determine the conformation and molecular orientation that finally yields the most
favorable score between each small molecule and a protein target [Seifert and Lang, 2008].

Turk et al., [2013] successfully performed virtual screening against a Mitis group of *Streptococcus*. They found a new hit compound which with micromolar quantities was identified to inhibit MurF enzymes from *S. pneumoniae*. Recently, RNase J1 and J2 were found to be essential for the growth of group a *streptococcus*. Virtual screening was performed, later to which cell-based biological assay identified two compounds with good activity [Hu et al., 2014]. Such application of virtual screening has also been made in other antibiotic resistant bacteria too. A diverse combinatorial library of amino-oxazole derivatives was designed by Brylinski and Waldrop [2014], to inhibit bacterial biotin carboxylase which is involved in fatty acid synthesis of bacteria.

Watowich and his co-workers [2014], have newly constructed web-based portal (DrugDiscovery@TACC) for structure-based virtual screening. This is an extension of virtual screening studies performed using IBM's World Community Grid, facilitated access to supercomputer resources managed by the Texas Advanced Computing Center (TACC). The server includes druglike commercially available small-molecule libraries. This server uses AutoDock Vina to perform individual docking analysis. The reliability of AutoDock has been proven by playing an early role in developing the first clinically approved inhibitor for HIV integrase [Schames et al., 2004]. Using the power of grid computing in combination with AutoDock Vina, a novel inhibitor of Dengue virus protease was identified by the establishers of DrugDiscovery@TACC [Viswanathan et al., 2014].