

## Emerging trends at the interface of chemistry and biology: Applications to the design of human therapeutics\*

SANTANU BHATTACHARYA<sup>1,\*\*</sup> and RAGHAVAN VARADARAJAN<sup>2</sup>

<sup>1</sup>Department of Organic Chemistry, Indian Institute of Science, Bangalore 560 012

<sup>2</sup>Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012

e-mail: sb@orgchem.iisc.ernet.in; varadar@mbu.iisc.ernet.in

**Abstract.** This article describes recent developments in the design and implementation of various strategies towards the development of novel therapeutics using first principles from biology and chemistry. Strategies for multi-target therapeutics and network analysis with a focus on cancer and HIV are discussed. Methods for gene and siRNA delivery are presented along with challenges and opportunities for siRNA therapeutics. Advances in protein design methodology and screening are described, with a focus on their application to the design of antibody based therapeutics. Future advances in this area relevant to vaccine design are also mentioned.

**Keywords.** Multi-target therapeutics; network biology; gene delivery; protein therapeutics; protein design.

### 1. Introduction

Over the past few decades, there have been significant advances in the fields of chemistry, molecular, structural and cell biology, and genetics, which have led to an increased understanding of the various life processes in general. However, much remains to be explored. Rapid increases in human population, overcrowding, environmental degradation and rapid transport have all enhanced the spread of existing infectious diseases as well as the threat of newly emerging infections. At the same time increased mechanization, sedentary lifestyles, dietary changes and an increase in life expectancy have led to increased rates of a variety of non-infectious diseases. Advances in diverse areas of chemistry and biology are being used to discover therapeutics for both infectious and non-infectious diseases. Novel strategies have been discovered to design and make new active principles. This article discusses a few of them, which appear to be promising. The article is divided into two parts; the first part describes pro-

gress in the area of small molecule therapeutics, gene delivery and siRNA technology. The next part describes recent progress in the area of protein therapeutics.

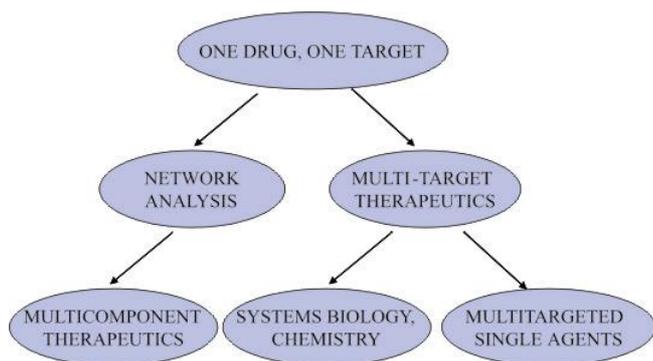
Organic chemistry is an enabling principle for the development of molecules that can selectively perturb or probe various biological systems. There is a growing understanding of the identification of molecules or systems that engage in distinct interaction with one or several components of complex and dynamic biological milieu. Here we present instances where chemical biology principles have been extended for different therapeutic leads.

### 2. Multi-target therapeutics and network analysis

The modern drug discovery has shifted its focus from the concept of one gene, one drug, and one disease, i.e. single-target strategy to multitarget strategy, which is based on network biology (figure 1).<sup>1,2</sup> Former suffers from lack of efficacy, toxicity, drug resistance and huge investment. Network biology is a tool to overcome this, which emphasizes on tackling the elements that disturb the diseased network rather than targeting the specific gene causing the disease.

\*Reproduced from 'Current Trends in Science', Platinum Jubilee Special Publication, Indian Academy of Sciences, Bangalore, 2009. pp. 9–19.

\*\*For correspondence



**Figure 1.** Drug development strategies.

### 3. Strategies of multi-target therapeutics

(a) Multiple drugs are given in combination, as in antiretroviral therapy for AIDS. But patient compliance and adverse drug-drug interactions are the main drawbacks of this method.

(b) Multicomponent therapeutics where two or more active ingredients are effectively combined into a single delivery system and administered as a drug. It is easy to find the optimal formulation with different agents that achieve maximum potency. If required, the target can be subjected to differential exposure of the individual agents by a programmed release of the constituting agents. However, differences in pharmacokinetics, bioavailability and metabolism of the individual ingredients cause many problems. Examples include, drugs e.g. Atripla (HIV drug), Advair (a drug for chronic obstructive pulmonary diseases), Caduet (used for treatment of high blood pressure and high cholesterol), etc.

For either of the above two-drug combination strategies, not only should the individual drug be safe but should be so also in combination. This limits the availability of safe drug combinations.

(c) Individual multi-targeted single agents, which use the principle of polypharmacology, where a single compound specifically binds to two or more molecular targets.<sup>3</sup> Such a drug circumvents the problems regarding equitable pharmacokinetics and biodistribution. The formulation of an individual active agent is easier compared with that of a mixture. Examples of multi target drugs are: Symbalta (cymbalta), Sutent (an oral multikinase inhibitor, an anti-cancer drug for multiple cancer), Nexavar (an oral multi-kinase inhibitor, an anti-cancer drug for liver and kidney), etc. Cymbalta is an antidepressant drug belonging to serotonin and norepinephrin reuptake inhibitors, which works by restoring their

balance in the brain. It is used to treat depression and anxiety disorder.

This is also used for managing pain caused by nerve damage in patients suffering from diabetic neuropathy.

Now to make polypharmacological drugs, one has to identify a lead compound which has activity against multiple targets without becoming nonselective and then this should be transformed into a drug with a good pharmaceutical profile. More challenging is to optimize the potency at both the targets. To predict a multi-target compound requires high level computational methods to explore a vast number of possible drug and target combinations which are compatible with each other.

### 4. Multi-target therapeutics and network analysis for cancer treatment

Most of the current anticancer drugs have low therapeutic index because their action is based on inhibiting the essential functions that are present in both normal and cancerous cells. Their effectiveness is because of rapidly dividing nature of the cancer cells compared to the normal cells. It would be ideal if anticancer agents target proteins or interactions that are essential in cancer cells but non-essential in normal cells. The genetic changes in a cancer cell change the relative dependence on specific proteins relative to normal cells. Thus the new methodology takes advantage of this difference in topology of the biological network to produce drugs with better therapeutic index.

#### 4.1 Drug design based on network analysis

Effect of a drug is studied in the context of relevant interactions (link) present in protein, regulatory, metabolic or signalling networks where the respective target is called element. Analysis based on genomic, proteomic, metabolism, etc. are the methods to identify the systemic analysis of multitarget drug action. A simple topological model is created by suppressing an element or a link and then its effect on the network is studied. It is found that multiple but partial attacks on various elements or links is more efficient than complete attack on a single element or link because the former affect more number of network links.<sup>4</sup> Thus, multi-target drugs are better than single target drugs because they affect more sites of network. The effect is more if they are distributed over the entire network.

#### 4.2 Mutations in phosphatidylinositol 3-kinase – a cause for cancer

Phosphatidylinositol 3-kinases/Phosphoinositide 3-kinases (PI 3-kinases or PI3Ks) are enzymes which can phosphorylate the position-3 hydroxyl group of inositol ring of phosphatidylinositol. Discovery of cancer specific mutations in the catalytic subunit p110 $\alpha$  of class I **PI3K** has made it a promising cancer target.<sup>5</sup>

The p110 $\alpha$  mutations show a preference to three mutational hot spots within the coding region of PIK3CA. Many of the mutations in p110 $\alpha$  due to cancer are missense mutations where an amino acid is substituted with the change in a single nucleotide. The p110 $\alpha$  hot spot mutants can grow independently, resistant to apoptosis and can spread the tumour elsewhere.<sup>6</sup>

Many other rare mutations accompany p110 $\alpha$  mutations with some gain of function like activating growth factor-independent phosphorylation of Akt and p70s6K and enhancing the lipid kinase activity. Nonetheless rare types of mutants are less efficient than the hot-spot mutants with respect to initiation or replication of the cancerous cells.

Based on molecular modelling, function of mutations present in various functional domains is proposed. Homology model proposes that the gain of function in p110 $\alpha$  might involve cellular localization, interaction with other proteins and modification of the catalytic activity.

#### 4.3 Mutants in p110 $\alpha$ as cancer targets

The drug should be isoform specific because  $\alpha$ -isoform is the main regulator of the cell growth.<sup>7</sup> Small molecule inhibitors are good candidates because they can be cancer specific and PI3K being an enzyme can be targeted easily by them. The presence of mutational hot spots can be taken into advantage to achieve mutant specificity. It is found that substitution of H1047R creates most robust cancerous p110 $\alpha$  with highest frequency. Now a compound that can differentiate between the mutant and wild-type protein, thus binding with two adjacent sites on the target protein, can be a promising candidate.

A deeper understanding of the structure of wild-type and mutant p110 $\alpha$  would greatly enhance the identification of mutant-specific small molecule inhibitors and throw light on how mutant-mediated gain of function takes place.

### 5. Challenges in the development of HIV medicine

Most dreaded disease of the modern day world is AIDS (Acquired Immuno Deficiency Syndrome), which is caused by HIV-1 virus. A vaccine for HIV-1 seems to be still elusive for the scientists because the virus is highly diverse in nature, and escapes the adaptive immune responses, difficult to induce broad reactive antibody responses and its viral reservoirs are set up very early.<sup>8</sup> Though there are drugs which partially prevent the disease by controlling the viral replication to a certain extent, there is no vaccine which completely blocks the infection and provides full immunity.

#### 5.1 Life cycle of HIV

The viral RNA and other contents enter into the cytoplasm of the cell after the virus attaches itself to specific receptors in lymphocyte membrane. The enzyme reverse transcriptase transcribes the viral RNA sequence into a complementary DNA sequence and the enzyme 'integrase' coalesces with host cell DNA. Replication of viral RNA is triggered by activation of lymphocyte which is followed by translation to produce various structural proteins like reverse transcriptase, protease, integrase, etc. Proliferation occurs when infected proteins bind to healthy immune cells.<sup>9</sup>

### 6. Viral entry inhibitors

(a) Enfuvirtide (DP-178, T-20), a peptide inhibits HIV fusion by preventing the fusion of a glycoprotein (gp41) of the viral envelope with the cell membrane. This is taken along with other antiretrovirals but its low genetic barrier to resistance undermines its usefulness.

(b) Small molecule antagonists like mozobil, maraviroc, vicriviroc and aplaviroc hinder the viral entry by targeting human chemokine receptors CCR5 and CXCR4.

#### 6.1 HIV protease inhibitors

HIV protease belongs to an aspartic protease family which plays an indispensable role in the life cycle of HIV. Protease inhibitors block protease activity by mimicking the structure of its regular substrate and subsequent competitive binding with the virus.

Some examples are atazanavir, tipranavir and darunavir, etc.

### 6.2 HIV reverse transcriptase inhibitors

These are of two types – nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). NRTIs are first converted into their active metabolites and then they act as competitive inhibitors or alternative substrates with respect to the normal substrates (dATP, dGTP, dCTP or dTTP) and lead to the termination of chain elongation. Some examples are zidovudine, didanosine, zalcitabine, etc. Nucleotide reverse transcriptase inhibitors (NtRTIs) also have similar mode of action but they differ in the number of steps required for their activation.

### 6.3 HIV non-nucleoside reverse transcriptase inhibitors

They block the enzyme's activity by interacting with an allosteric binding site on HIV-1 reverse transcriptase. Some examples of this category include nevirapine, delavirdine, etc.

## 7. Chemical interface in lipid rafts

Stier and Sackmann postulated the concept of membrane microdomains in 1970 (ref. 10) which was later supported and further studied in detail by various other groups. Gerrit van Meer coined the term 'lipid rafts' for these microdomains which are enriched with lipids, cholesterol, glycolipids and sphingolipids present in cell membranes.<sup>11</sup> Although this concept was invoked to explain the transport of cholesterol from the trans golgi network to plasma membrane, recent research<sup>12</sup> has shown that these lipid rafts have roles in diverse processes like signal transduction, endocytosis apart from cholesterol trafficking.

Composition of lipid rafts can be a well-ordered, cholesterol- and glycosphingolipid-enriched domain or a diverse collection of lipids, showing variable enrichment in cholesterol and sphingolipids depending on the method of preparation. Lipid rafts are related to the immiscibility of ordered and disordered liquid microdomains in cells which cause decrease in free energy between two phases. Membrane proteins are categorized based on their loca-

tion whether they are present in rafts or in liquid-disordered phase or in the intermediate state. Thus, lipid rafts constitute proteins and lipids that not only float freely within the liquid disordered bilayer of cellular membranes but can also aggregate to form clusters. Removal of cholesterol which partitions between raft and non-raft phase makes the proteins non-functional.

Lipid rafts are also implicated in many viral and bacterial infections.<sup>13</sup> For example, influenza virus and HIV-1 virus enter into the cell by binding at the rafts followed by altering the cell functions. In diseases like HIV and leishmaniasis, pathogens interfere with the function of rafts of the host cell and use them to gain entry into the cell interior.<sup>14</sup> Apart from this the size of lipid raft is very important in raft-borne signaling proteins. They are in 'off' state when rafts are small and turned on when rafts aggregate to form clusters where functionally related proteins interact and signal transduction occurs. The size of a lipid raft can be anywhere between 10 nm to 1000 nm and they always restructure themselves by continuous forming and breaking. So it is quite difficult to study their evolution.

## 8. Gene delivery

According to Miller, gene therapy is defined either as 'the use of genes as medicines to treat disease' or 'the delivery of nucleic acid (with a vector) to patients for some therapeutic purpose.'<sup>15</sup> In other words, it involves the delivery of the DNA to the targeted cells, thus curing the disease at its causative level. For this to be successful, an efficient delivery system should be available which transports the gene expression system into the cell that encodes a therapeutic protein that is either not produced in the patient or produced in a nonfunctional form.

The fact that naked DNA is not suitable for the *in vivo* transport of genetic materials into selected cell body types due to its degradation by serum nucleases, search for efficient carrier system has intensified over the past few years. Vectors can be divided into two classes: viral or non-viral. Some of the viral vectors are adenovirus, retrovirus, adeno-associated virus, etc. However immunogenicity, toxicity, insertional mutagenesis and vector inactivation, etc. have led the scientists to develop nonviral vectors.

Among non-viral vectors, DEAE-dextran mediated transfection and calcium-phosphate methods are classical and popular methods. Now the research

is focused on the design and synthesis of, cationic lipids,<sup>16</sup> cationic polymers<sup>17</sup> and peptides which are highly efficient gene transfection vectors. Cationic liposomes or polymers bind to negatively charged DNA due to electrostatic interactions and form a complex. This binds to the cell surface and endocytosis takes over. But after entering lysosome, there is no clear understanding as to how the release of DNA takes place before the occurrence of lysosomal degradation although the addition of buffering agents can reduce this to some extent.

Cationic lipids are converted into liposomes which contain inner aqueous pools capable of encapsulating drugs or DNA.<sup>16</sup> Lipoplexes thus formed gain entry into the cell by slow endocytosis. The nucleic acids which escape from endosomal compartments enter into the cell cytoplasm. Nucleic acids may face further difficulty in entering the nucleus. Once it enters the nucleus, expression of delivered genes occurs.

Typically, a cationic lipid consists of a hydrophobic domain (usually two long aliphatic chains or cholesterol-type group), a cationic head group and a linker which connects these two parts. Many of the cationic lipids have glycerol backbone on which the hydrophobic tails are present.<sup>18</sup> Even the head group is attached through a hydrophilic domain like polyethylene glycol to impart special properties to the lipid. The systematic modification of each part and understanding of the structure activity relationship is an active area of research by many groups.<sup>19,20</sup>

There are many advantages of gene delivery by cationic lipids. Lipoplexes are easy to produce, they are non-immunogenic, can deliver large polynucleotides into the cell. Ease of synthesis allows one to incorporate various functionalities to fine tune the efficiency. However, poor cell specific binding, low cellular uptake of plasmid DNA, serum incompatibility and toxicity are some of the issues that need to be addressed carefully in order to obtain an efficient cationic lipid for gene transfection.

Small interfering RNA (siRNA), sometimes known as short interfering RNA or silencing RNA, is a class of 20–25 nucleotide-long double stranded molecules that play a variety of roles in biology.<sup>21</sup> Most notably, siRNA is involved in the RNA interference (RNAi) pathway, where it interferes with the expression of a specific gene. It was observed that when double-stranded RNAs (ds-RNA) are injected into the worm *Caenorhabditis elegans*, genes, whose sequences were complementary to those of the introduced ds-RNAs were silenced.<sup>22</sup> Now it is

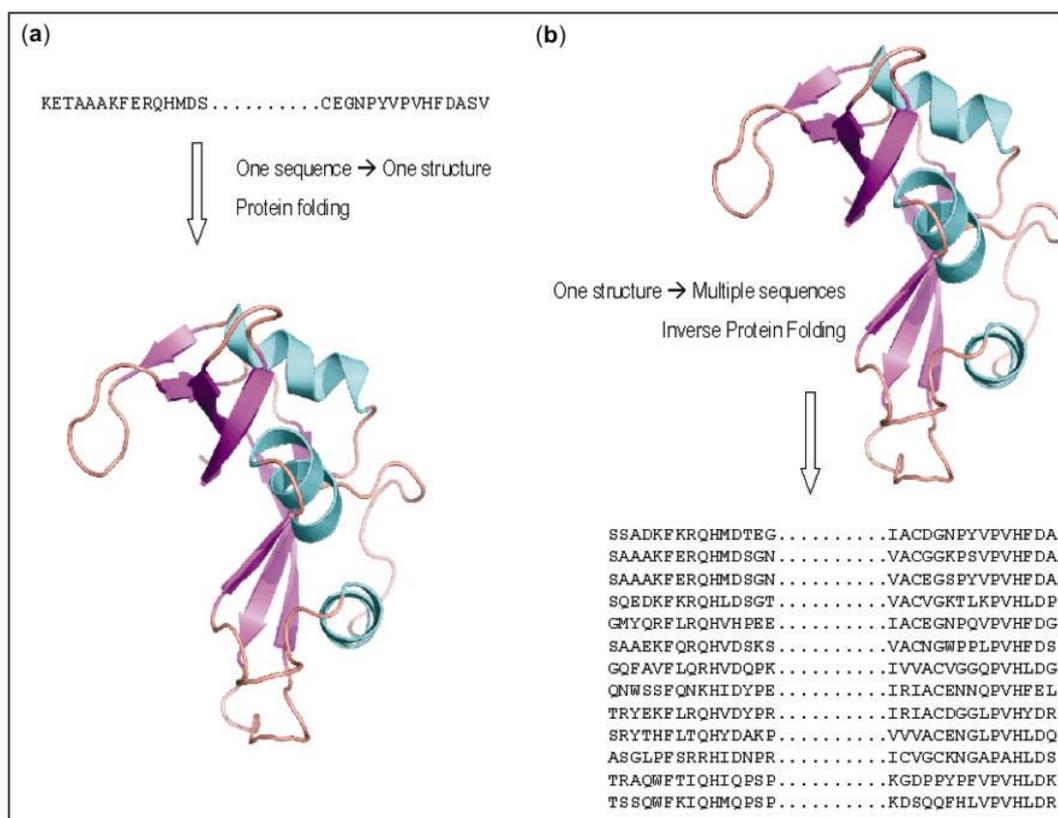
known that ds-RNAs are processed into short interfering RNAs (siRNAs) by the RNase enzyme Dicer. These siRNAs are then incorporated into a silencing complex called RISC (RNA-induced silencing complex), which identifies and silences complementary messenger RNAs.<sup>23</sup> Delivery of siRNA directly in cells can be achieved by using microinjection, electroporation or by the use of a transfection reagent.

In addition to their role in the RNAi pathway, siRNAs also act in RNAi-related pathways, e.g. as an antiviral mechanism or in shaping the chromatin structure of a genome. Also it has been shown that ds-RNAs targeting gene promoters induce potent transcriptional activation of associated genes and these are termed small activating RNAs (saRNAs).

The fact that RNAi via siRNAs can knock down any gene of interest, it has a great potential to contribute to modern medicine.<sup>24</sup> Because disease processes also depend on the activity of multiple genes, it is expected that in some situations turning off the activity of a gene with a siRNA could produce a therapeutic benefit. There is also potential for using RNAi for the treatment of viral diseases such as those caused by the hepatitis C virus (HCV) and the human immunodeficiency virus (HIV). Synthetic siRNAs and expressed shRNAs have been used to target several early and late HIV-encoded RNAs in cell lines and in primary haematopoietic cells. When a hepatitis B virus (HBV) was co-delivered with an expression unit encoding an anti-HBV shRNA in mice, a significant knockdown (99%) of the HBV core antigen in liver hepatocytes could be achieved. Many studies have used siRNAs as an experimental tool to dissect the cellular pathways that lead to uncontrolled cell proliferation and to cancer. Thus RNAi has been proposed as a potential treatment for cancer.

## 9. Challenges using siRNA technology

Sometimes non-specific effects are triggered by the experimental introduction of a siRNA. A mammalian cell may mistake siRNA as a viral by-product and mount an immune response. Off-targeting can also occur where the genes with incomplete complementarity (microRNAs) are inadvertently down-regulated by the siRNA and cause misinterpretation of data and toxicity. Introduction of too much siRNA can result in non-specific events due to activation of intrinsic immune responses. Thus intracellular delivery at therapeutically effective concentrations is yet to be achieved.



**Figure 2.** Schematic description of protein folding and inverse folding problems. (a) The protein folding problem deals with prediction of a protein three-dimensional structure from its amino acid sequence. (b) The inverse folding problem is to predict all possible sequences compatible with a given protein three-dimensional structure.

To derive maximum benefit from this emerging technology, strategies that maintain a high level of siRNA efficiency while minimizing the non-specific off-target effects associated with siRNA expression should be explored in detail.

## 10. Protein therapeutics

### 10.1 *Folding and inverse folding problems*

Proteins are biopolymers that are typically composed of 20 different kinds of monomer units (amino acids) linked by peptide bonds. Most proteins fold into a well-defined three-dimensional structure. The structure of a protein is intimately linked to its function. For many proteins, the information specifying the structure is encoded in the amino acid sequence. This can be shown by denaturing purified protein in aqueous solution and allowing the protein to refold by removal of the denaturing agent. Although it is widely accepted that sequence specifies structure, it is still not possible to predict

the structure of a protein from first principles given its amino acid sequence. This problem is known as the protein folding problem and it remains one of the major unsolved problems in biology. A related problem, known as the inverse folding problem has to do with enumeration of all possible sequences consistent with a given protein structure (figure 2). Solutions to this problem have important implications for protein design. Below we briefly describe progress in this area and applications to the design of biological therapeutic agents.

In proteins, bond lengths and bond angles between different types of atoms are highly conserved and depend only on the type of atom(s) involved, not the specific amino acid or protein. Differences in three dimensional structures between proteins arise primarily from differences in dihedral angles. This insight was made use of by Ramachandran and colleagues<sup>25</sup> who showed that protein main chain configurations are described by two dihedral angles ( $\phi$  and  $\psi$ ) and that substantial regions of conformational space are disallowed because of steric overlap

between non-bonded atoms. While Ramachandran and colleagues focused on steric restrictions imposed primarily by the main chain, protein structures are also constrained by side chain interactions. A major advance in addressing the inverse folding problem came from the work of Ponder and Richards.<sup>26</sup> They demonstrated that approximating side chain conformations by discrete rotamers from a rotamer library, combined with a few additional constraints (avoidance of steric overlap, buried charge and large cavities) greatly restricted the number of sequences energetically compatible with a given main chain configuration/fold. One of the major challenges in inverse folding or protein design is addressing the combinatorial nature of the problem. For example, for a 100 residue protein, there are 20 possible amino acids at each position. If we assume that each amino acid on an average can be modeled by three rotameric configurations, there are still  $(20 \times 3)^{100}$  i.e.  $\sim 10^{177}$  possible configurations that need to be sampled. A number of elegant approaches and approximations have been used<sup>27,28</sup> to reduce the search space and provide approximate solutions to the problem. A more tractable and commonly encountered problem is the structure based prediction of a limited number of mutations to alter specific properties of a protein. For example, mutations to change protein stability or to modulate protein: protein or protein: substrate interactions. Major limitations in protein design arise from errors in the potential function(s) employed and the use of various approximations and cut-offs to compute energies of specific sequence–structure combinations. It is therefore essential to be able to experimentally construct and test a large number of protein sequences and then experimentally select the optimal ones from a large pool. There are a variety of methodologies for doing this, which include phage display, yeast display and ribosome display.<sup>29</sup>

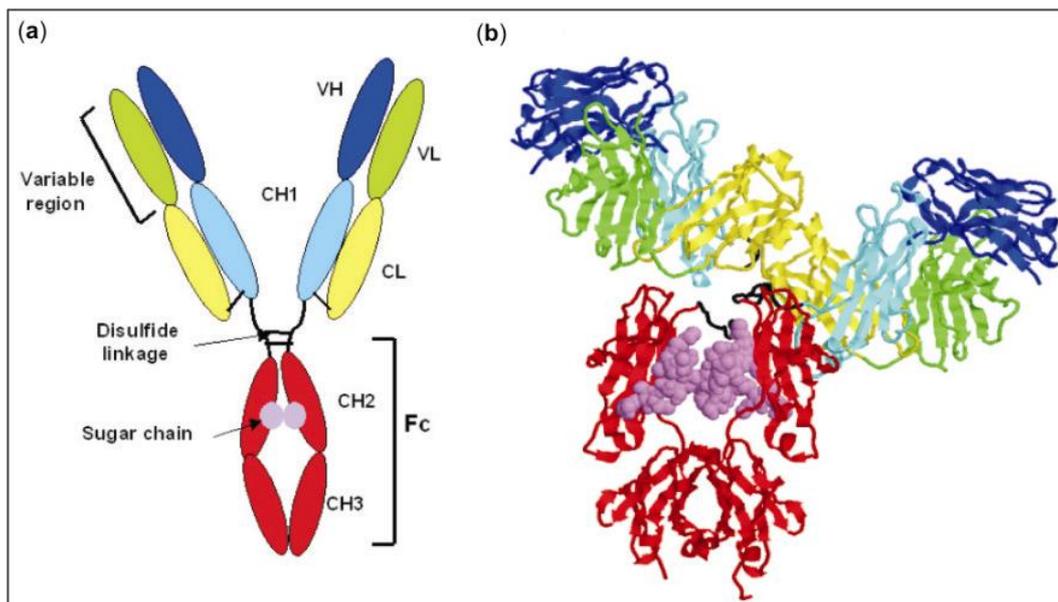
## 11. Antibodies as therapeutics

The vast majority of therapeutics currently used to treat disease are small molecules with molecular weights of  $< 500$  Da.<sup>30</sup> In recent years there have been increasing numbers of ‘biological’ therapeutics developed by the biotechnology industry. Monoclonal antibodies (mAbs) and related molecules constitute the major fraction of protein therapeutics with 2006 US sales of \$14.5 billion<sup>31</sup> and an annual growth rate  $> 35\%$ . However, the high cost of these

treatments (of the order of Rs 100,000 per dose in some cases), have put them out of reach of the vast majority of people, especially in the developing world. It is therefore imperative to develop better and cheaper ways of developing and producing these molecules. Antibodies (Abs) typically produced by the body in response to infection by a foreign agent are a complex mixture of molecular species, termed polyclonal antibodies. However, in 1975 Milstein and Kohler<sup>32</sup> showed that large amounts of homogeneous antibody can be produced by fusion of an Ab producing cell with a myeloma cell. Since the Ab producing cell is typically from a mouse that has been immunized with the target protein (antigen), the resulting Abs need to be modified (humanized) before they can be used for therapeutic purposes in humans, as mouse derived Abs will provoke an immune response in humans. However, mouse derived Abs can be used without modification in diagnostic applications. Other drawbacks of conventional hybridoma technology are that Abs cannot be raised against self molecules, toxins and some small molecules because of toxicity and/or lack of immunogenicity. An alternative is the use of *in vitro* technology such as phage display or yeast surface display.<sup>29</sup>

## 12. Antibody structure

A schematic description of an immunoglobulin G (IgG) type Ab molecule is shown in figure 3. These constitute the major type of Ab in serum. The molecule consists of a total of four polypeptide chains, two identical pairs of light chains and two identical pairs of heavy chains. Each light chain interacts with a single heavy chain through disulfide bonds and noncovalent interactions. The N terminal  $\sim 110$  amino acids of each chain constitute the variable regions. Within the variable regions, maximal sequence diversity occurs within the so called complementarity determining regions (CDRs) which are directly involved in antigen binding. The CH<sub>2</sub> region contains a structurally and functionally important conserved glycosylation site. In addition, there may be other nonconserved glycosylation sites in other regions of the molecule. The constant region (Fc) of an Ab is involved in effector functions which involve binding to receptors (Fc receptors) on immune cells, such as natural killer (NK) cells or macrophages. Binding of multiple Abs to an antigen *in vitro* results in clustering of Fc regions and binding



**Figure 3.** Antibody structure. (a) Schematic diagram of an immunoglobulin G type of Ab. Variable and constant domains are prefixed with V and C, respectively. The VH and VL variable regions are involved in antigen binding. The Fc region involved in effector function consists of the CH<sub>2</sub> and CH<sub>3</sub> domains. (b) Structure of mAb b12 (adapted from the pdbid 1HZH) using the program RASMOL (<http://www.rasmol.org>). Colours of the corresponding regions of (a) and (b) are identical.

to Fc receptors, which in turn targets cells or particles displaying the antigen for destruction. In the following sections, we describe approaches that have been used to design and select therapeutic antibodies (Abs) for use in humans and the possible use of similar technology for vaccine development. The first mAbs to be tested in humans were rodent derived Abs developed by conventional hybridoma technology. However these suffered from several problems, such as low half-lives in humans, poor effector function and high immunogenicity (as they were recognized as foreign objects by the human immune system). More recently, these problems have been substantially addressed by the use of chimeric as well as humanized mAbs.<sup>33</sup> Chimeric mAbs are ones in which the VH and VL chains are murine (mouse) derived, while the constant regions are of human origin. Such mAbs contain approximately 33% of murine derived amino acid sequence. Humanized mAbs contain only about 10% of murine sequence primarily in the hypervariable loops. Recent methodologies such as phage display of Ab libraries or transgenic mice with human immunoglobulin genes<sup>34,35</sup> have made it possible to produce fully human mAbs (containing only human derived sequence) with desired specificities. It still

remains to be shown whether fully human mAbs provide significant therapeutic benefit over chimeric or humanized mAbs.<sup>36</sup>

### 13. Designer antibodies

Protein design and selection methodology has been extensively used to modulate antigen binding, effector function and the *in vitro* half lives of mAbs and to design Ab fragments that retain one or more functions of the intact molecule. The procedure for the *in vitro* selection of Abs from large libraries is summarized in figure 4. For mAb libraries generated *in vitro*, the diversity of the library can be generated either from synthetic variable genes or using variable regions cloned from human B cells. During *in vitro* selections, the phenotype of the mAb displayed on a particle is linked to the genotype (DNA sequence encoding the mAb/fragment) which is packaged inside the particle (figure 4). The library consists of a large number of such particles (bacteriophages, viruses or cells). In one round of selection, the library is bound to immobilized antigen and unbound library members are washed off. Bound members are eluted, allowed to replicate in an appropriate host, re-purified and subjected to further

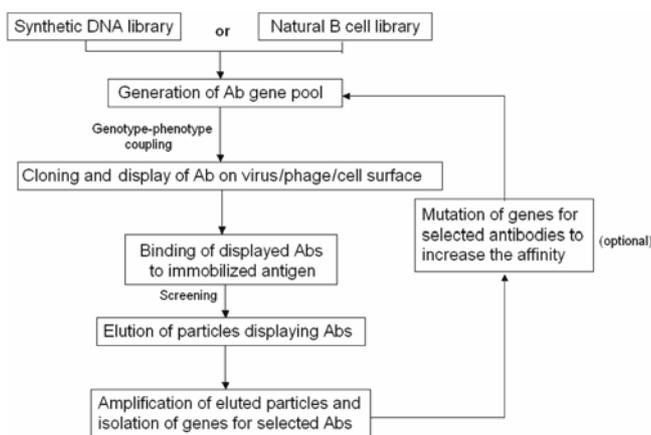
rounds of selection. At each round, the fraction of the input pool that binds is measured. Typically after a few rounds of selection, a significant fraction of the input pool shows binding. At this point individual members are sequenced, expressed, purified and characterized for binding affinity to the desired target. The diversity of libraries that can be screened *in vitro* is typically limited by the transformation efficiency of the host cell which is typically less than  $10^9$  transformants/ $\mu\text{g}$  of DNA. Since there are 20 possible amino acids at each position, even a library where only 8 positions are randomized will have a diversity ( $20^8 = 256 \times 10^8$ ) in excess of what can be conveniently transformed. It is therefore important to use computational analyses to guide the choice of regions to be mutated as well as to introduce specific types of diversity designed to simulate the types of amino acids that are abundant in natural binding sites.<sup>37</sup> Recent studies have attempted to define the minimum diversity in amino acid sequence required for efficient recognition of antigens by mAbs. Eleven synthetic libraries were constructed with CDR diversity restricted to only four amino acids (including at least one small amino acid to provide conformational flexibility) and screened for binding to four different protein antigens. Two libraries (with amino acids Y/A/D/S and H/R/D/G) generated Abs against 3 of the 4 antigens.<sup>38</sup> In another study<sup>39</sup> it was shown that functional Abs could be obtained with a diversity in CDR's generated from only two amino acids (Y and S). High affinity ( $K_d = 60$  nM) and high specificity Abs against human vascular endothelial growth factor were obtained even with such a binary recognition code.

In addition, to improving antigen binding affinity, protein design and selection strategies have also been used to improve Ab effector function. In a recent study,<sup>40</sup> automated protein design methodology combined with high throughput screening was used to engineer mutations in the Fc region of an anti-CD52 mAb. The best of the designed variants showed an increase in effector function of over 100 fold because of increased binding to the Fc $\gamma$  RIIIa receptor. In another study<sup>41</sup> the *in vivo* half of Abs and Ab fragments was modulated by mutations in the Fc region to alter interactions with the protective neonatal Fc receptor FcRn. This has both therapeutic and imaging applications.

One major barrier to the use of mAbs as therapeutics is high cost. In part this is because they need to be expressed using expensive mammalian cell culture systems to ensure correct glycosylation. Recently it has been possible to engineer yeast and other microbes to produce glycosylation patterns similar to those found in human cells<sup>42–45</sup>. These advances have the potential to dramatically lower costs of mAbs by expression in microbial systems. Alternative methods that can potentially lower costs of treatment are the use of genetically engineered cells for the continuous product of Abs *in vivo*.<sup>46</sup>

#### 14. Vaccine design

As can be seen from the above discussion, the development of Ab therapeutics has been facilitated by a combination of developments in computational methodology for protein design, library construction, screening via phage/yeast display, development of appropriate transgenic mice and advances in understanding of factors modulating Ab efficiency *in vitro*. All the tools used in Ab design can also be applied to design better vaccines. In the developing world, infectious diseases are the leading cause of death. For three of the major infectious diseases (TB, malaria and HIV) no vaccine exists. For other diseases such as influenza, a satisfactory vaccine currently exists. However, the advent of highly pathogenic avian influenza (bird flu) and pandemic H1N1 is a reminder that new strategies to counter such newly emerging infections are urgently required. New approaches to vaccine design involve identification of the target antigen, followed by determination of its molecular structure as well as mapping regions of the structure (neutralization epitopes) where Ab binding will reduce or abrogate in-



**Figure 4.** Scheme for *in vitro* selection of Ab/Ab fragments that bind to a given antigen.

fectivity/pathogenicity. Following this, protein design methodology can be used to construct fragments/domains of the protein containing the desired neutralization epitopes. The designed proteins can be screened for binding to known neutralizing mAbs. Selected molecules (immunogens) can be injected in appropriate animal models to see if they can induce desired Ab responses and, where feasible, protect the animals from pathogenic challenge. Immunogens which show encouraging responses in small animals can be further tested in non-human primates and eventually in humans. This approach, known as 'reverse vaccinology' or 'retrovaccinology'<sup>47</sup> has the potential to accelerate the development of vaccines for emerging infections. It can also be used to reduce costs of existing vaccines, by re-designing immunogens so that they can be expressed in microbial systems instead of currently used more expensive cell culture systems. However, further work needs to be done to assess the benefits of this methodology over conventional methods of using attenuated or killed versions of the infectious organism as vaccines.

## 15. Conclusions

Multitarget therapeutics as well as network analysis derived multicomponent therapeutics have the potential to treat diseases that have proved resistant to conventional single target therapeutics. Gene therapy and siRNA are alternatives to conventional drug treatment. Advances in gene delivery via cationic lipids are important for the enormous potential benefits of gene therapy to be realized. Small molecule drugs are by far, the most common allopathic therapeutics. However, the market for biological therapeutic proteins, especially antibodies and antibody derivatives is rapidly growing. Advances in computational biology, structural biology and immunology have led to the introduction of numerous mAb therapeutics in recent years. Improvements in protein design and screening methodologies have made it possible to modulate the stability, *in vivo* half-life and function of mAbs and similar molecules in a semi rational fashion. Similar methodologies can potentially be applied to design immunogens and vaccines for newly emerging global health threats.

## Acknowledgements

S B thanks V Guru Raja for his help with the manuscript organization. R V thanks Bharat Adkar and

Piyali Saha for providing figures and Bhagyashree K G for help with manuscript preparation.

## References

1. Russell R B and Aloy P 2008 *Nat. Chem. Biol.* **4** 666
2. Zimmermann G R, Lehar J and Keith C T 2007 *Drug Discov. Today* **12** 34
3. Hopkins A L 2008 *Nat. Chem. Biol.* **4** 682
4. Whitehurst A W, Bodemann B O, Cardenas J, Ferguson D, Girard L, Peyton M, Minna J D, Michnoff C, Hao W, Roth M G, *et al* 2007 *Nature* **446** 815
5. Engelman J A, Luo J and Cantley L C 2006 *Nat. Rev. Genet.* **7** 606
6. Hennessy B T, Smith D L, Ram P T, Lu Y and Mills G B 2005 *Nat. Rev. Drug Discov.* **4** 988
7. Knight Z A, Gonzalez B, Feldman M E, Zunder E R, Goldenberg D D, Williams O, Loewith R, Stokoe D, Balla A, Toth B *et al* 2006 *Cell* **125** 733
8. Barouch D H 2008 *Nature* **455** 613
9. De Clercq E 2007 *Nat. Rev. Drug Discov.* **6** 1001
10. Stier A and Sackmann E 1973 *Biochim. Biophys. Acta.* **311** 400
11. Simons K and van Meer G 1988 *Biochemistry* **27** 6197
12. Simons K and Toomre D 2000 *Nat. Rev. Mol. Cell Biol.* **1** 31
13. Carter G C, Bernstone L, Sangani D, Bee J W, Harder T and James W 2009 *Virology* **386** 192
14. Garg R, Lodge R, Descoteaux A and Tremblay M J 2008 *J. Infect. Dis.* **197** 1701
15. Miller A D 1998 *Angew. Chem. Int. Ed.* **37** 1769
16. Bhattacharya S and Bajaj A 2005 *Curr. Opin. Chem. Biol.* **9** 647
17. Bajaj A, Kondaiah P and Bhattacharya S 2008 *Bioconjug. Chem.* **19** 1640
18. Bhattacharya S and De S 1999 *Chemistry-A European J* **5** 2335
19. Bajaj A, Kondaiah P and Bhattacharya S 2007 *Bioconjug. Chem.* **18** 1537
20. Bajaj A, Kondiah P and Bhattacharya S 2007 *J. Med. Chem.* **50** 2432
21. Hannon G J and Rossi J J 2004 *Nature* **431** 371
22. Fire A, Xu S, Montgomery M K, Kostas S A, Driver S E and Mello C C 1998 *Nature* **391** 806
23. Preall J B, He Z, Gorra J M and Sontheimer E J 2006 *Curr. Biol.* **16** 530
24. Gregory R I, Chendrimada T P, Cooch N and Shiekhattar R 2005 *Cell* **123** 631
25. Ramachandran G N, Ramakrishnan C and Sasisekhan V 1963 *J. Mol. Biol.* **7** 95
26. Ponder J W and Richards F M 1987 *J. Mol. Biol.* **193** 775
27. Lasters I, De Maeyer M and Desmet J 1995 *Protein Eng.* **8** 815
28. Kuhlman B and Baker D 2000 *Proc. Natl. Acad. Sci. USA* **97** 10383
29. Hoogenboom H R 2005 *Nat. Biotechnol.* **23** 1105

30. Lipinski C A, Lombardo F, Dominy B W and Feeney P J 2001 *Adv. Drug. Deliv. Rev.* **46** 3
31. Aggarwal S 2007 *Nat. Biotechnol.* **25** 1097
32. Milstein C 1980 *Sci. Am.* **243** 66
33. Lonberg N 2008 *Curr. Opin. Immunol.* **20** 450
34. Green L L, Hardy M C, Maynard-Currie C E, Tsuda H, Louie D M, Mendez M J, Abderrahim H, Noguchi M, Smith D H, Zeng Y *et al* 1994 *Nat. Genet.* **7** 13
35. Lonberg N, Taylor L D, Harding F A, Trounstein M, Higgins K M, Schramm S R, Kuo C C, Mashayekh R, Wymore K, McCabe J G, *et al* 1994 *Nature* **368** 856
36. Gibson T B, Ranganathan A and Grothey A 2006 *Clin. Colorectal Cancer* **6** 29
37. Sidhu S S, Li B, Chen Y, Fellouse F A, Eigenbrot C and Fuh G 2004 *J. Mol. Biol.* **338** 299
38. Fellouse F A, Wiesmann C and Sidhu S S 2004 *Proc. Natl. Acad. Sci. USA* **101** 12467
39. Fellouse F A, Li B, Compaan D M, Peden A A, Hymowitz S G and Sidhu S S 2005 *J. Mol. Biol.* **348** 1153
40. Lazar G A, Dang W, Karki S, Vafa O, Peng J S, Hyun L, Chan C, Chung H S, Eivazi A, Yoder S C, *et al* 2006 *Proc. Natl. Acad. Sci. USA* **103** 4005
41. Olafsen T, Kenanova V E and Wu A M 2006 *Nat. Protoc.* **1** 2048
42. Chiba Y and Jigami Y 2007 *Curr. Opin. Chem. Biol.* **11** 670
43. Hamilton S R, Davidson R C, Sethuraman N, Nett J H, Jiang Y, Rios S, Bobrowicz P, Stadheim T A, Li H, Choi B K, *et al* 2006 *Science* **313** 1441
44. Wildt S and Gerngross T U 2005 *Nat. Rev. Microbiol.* **3** 119
45. Kowarik M, Numao S, Feldman M F, Schulz B L, Callewaert N, Kiermaier E, Catrein I and Aebi M 2006 *Science* **314** 1148–1150
46. Luo X M, Maarschalk E, O'Connell R M, Wang P, Yang L and Baltimore D 2009 *Blood* **113** 1422
47. Burton D R 2002 *Nat. Rev. Immunol.* **2** 706