Cancer is one of the leading causes of death along with heart attack despite the enormous research and rapid development in the past decade. CD47 or integrin associated protein (IAP) is a 50 kDa widely expressed and highly glycosylated immunoglobulin transmembrane protein but its expression increases multifold in tumor cells. CD47 binds with SIRPα and participates in various activities responsible for regulation of apoptosis. It acts as a marker of self for cells so that the cells are not phagocytosed by the immune system. In vitro studies have shown that inhibition of CD47 helps in diminishing the tumor. Few antibodies are under Phase 1 clinical trials. But as it is known that antibody therapy is not cost effective so designing an organic anti CD47 drug molecule can be a good approach to utilize this target more efficiently. In this study we applied several computational approaches to identify new molecules as anti CD47 agents. Foldamers were designed using homologation and sequence based approaches and the binding characteristics of these molecules were studies using molecular docking and molecular dynamics studies. These molecules are presumed to be promising as potential anti cancer agents.
INTRODUCTION:
Cancer is the uncontrolled growth of the cells resulting into the formation of tumor. It is mainly of two types i.e. Benign cancer and malignant cancer. Benign cancer is the localized cancer, so it is not as much dangerous as the malignant cancer which is having metastatic property i.e. it can migrate from one tissue to another tissue present on a distant location in the body. Cancer continues to be one of the leading cause of death despite the enormous research and rapid development in the past decade.

By 2020 the total population of the world will cross 7.5 billion and 15 million new cancer cases will be diagnosed while 12 million cancer patients will die.

Cancer is caused by both internal (hormone, inherited mutation) as well as environmental factors (tobacco, diet, radiation). All of the cancers are the result of multiple mutations but most of these mutations are the result of the environmental interactions rather than genetically inherited. Hereditary cancers account for only 5-10% of total cancer cases. Others are the result of the environmental factors mainly smoking and alcohol consumption.

There are many therapeutic targets for different types of cancer like platelet derived growth factor, vascular endothelial growth factor, fibroblast growth factor and their respective tyrosine kinase inhibitors have become popular for the inhibition of the proliferation of endothelial cells while some cell cycle regulators like cell cyclin-dependent kinases, p21 gene and apoptosis modulators Bel-2 oncoprotein, p53 tumor suppressor gene, survivin protein etc are also some of the important targets. Some other proteins like farnesyltransferase (FTase), histone deacetylase (HDAC), and telomerase which play role in signal transduction are also attractive anticancer targets. And many drugs have been developed for these targets but they are not so much effective and their adverse effects are also very high. So, it has become the need of the time to consider an effective target like CD47 and prepare molecules against it.

CD47 or integrin associated protein (IAP) is a 50 kDa widely expressed and highly glycosylated immunoglobulin transmembrane protein. It is involved in many biological functions. It is encoded by CD47 gene located on chromosome 3. First of all it was found to be in association with the αvβ3 integrin but later erythrocytes were also found to express this protein. CD refers Cluster of Differentiation that means it is found in the cluster form. It is found on the surface of the leukocytes due to which it is also called Leukocyte surface antigen CD47.

CD in human is numbered from 1 to 371 upto April 2016. The number after CD (e.g. CD47) is allotted on the basis of the order of their discovery. They are actually the markers of the cells. The type of CD which is present on the leukocyte gives the idea of type of the cell (e.g. presence of CD45 and CD19 on the cell surface is the indication that it is a B-lymphocyte). Due to the specificity of the CD for each cell, it is of sufficient importance in immunophenotyping.

CD47 binds with SIRPα and participates in various activities responsible for regulation of apoptosis. It acts as a marker of self for cells so that the cells are not phagocytosed by the immune system. It is found to be actively involved in preventing the body cells against the attack of macrophages which are circulating throughout the body. In vitro studies have shown that inhibition of CD47 helps in diminishing the tumor. Few antibodies are under clinical trials. But as it is known that antibody therapy is not cost effective so designing an organic anti CD47 drug molecule can be a good approach to utilize this target more efficiently.
Structure of CD47

CD47 is a trans-membrane protein, it contains one extracellular IgV domain, five times trans-membrane spanning domains and one cytoplasmic tail. Structure-function studies have demonstrated that both the Ig domain and the MMS domain of CD47 are essential for its role in signal transduction as well as for localization of CD47 to the cholesterol-rich plasma membrane domains known as glycosphingolipid-enriched membranes (gems) or rafts. The Ig domain is required for association; the MMS domain apparently stabilizes association significantly. The mechanism of the MMS effect is likely to involve its ability to bind to cholesterol because cholesterol is required for stable, immune-precipitable association of CD47 with αvβ3.

Figure 1. Different leukocytes along with their specific CD surface antigen.3

Figure 2. Structure of CD47 (PDB ID: 2JJS, 1.8Å).

The cytoplasmic tail is of varying length from 4 to 36 amino acids. e.g. form 1 is consisting of 4 amino acids and it is found in epithelial and endothelial cells. Form 2 is the most common form
consisting of 16 amino acids and is highly expressed by monocytes, macrophages, neutrophils, basophils, eosinophils, dendritic cells, erythrocytes, megakaryocytes or platelets, T cells, B cells, natural killer cells, endothelial and epithelial cells; while the form 4 which contains 36 amino acids is primarily expressed in neurons, intestine, and testis. Mouse, rat and human show 70% similarity in their cytoplasmic tail sequence. The extracellular IgV domain is required for its binding to the ligand and the trans-membrane spanning domain stabilizes this interaction by binding to cholesterol because cholesterol is required for binding of CD47 to the αVβ3. By biochemical analysis it was found that the N-terminal residue of CD47 is glutamate which undergoes cyclization and converts into pyroglutamate. There are two short helical segments also. N-linked glycosation is identified at 4 out of 5 sites (N16, N32, N55 and N93). C23 and C96 which are available in the IgSF region are involved in disulphide linkage with trans-membrane residues which are found to be responsible for the stability of structure of the protein. The crystallographic data shows that the domains are present in the pairs and a strand swap also takes place between both the strands. But no evidence of the dimeric structure was found in gel filtration analysis and CD47/SIRPα structure also does not show any indications of the dimerism. Hence it was concluded that strand swap is not biologically relevant.4

SIRPα

SIRP family is a group of cell surface receptors which transmit the immune regulatory signal to the cell for optimal immune response. SIRPα, SIRPβ and SIRPγ are the main proteins of this family. Among them SIRPα and SIRPγ bind with CD47 efficiently rather than SIRPβ. In Jurkat T cells it is found that CD47 binding with SIRPα does not induce apoptosis while CD47 binding with SIRPγ induces apoptosis. SIRPα is mainly expressed by myeloid cells, neurons and stem cells. It is also a transmembrane protein like CD47.

The extracellular region contains three immunoglobulin superfamily domains- one V-set and two C1-set IgSF domain. Its cytoplasmic region is highly conserved among humans, rats and mice. The cytoplasmic tail contains various tyrosine(Y) residues which are phosphorylated when CD47 binds with SIRPα. It acts like Immunoreceptor Tyrosine-based Inhibitory Motif (ITIM). The enzymes of Src kinase family are responsible for this phosphorylation. After being phosphorylated, the other enzymes like phosphorytrosine phosphatases SHP-1 and SHP-2 are recruited which are responsible for the decrement of the activation of the molecules actively involved in cell signaling. It generates a ‘Do not eat me’ signal which inhibits the phagocytosis of the cell so it is said to be the “marker of self” whose expression prevents the cells from the attack of immune system. There are some anti SIRPα peptides which are found to inhibit the CD47- SIRPα interaction.6

Involvement of CD47 in various biological processes

Role of CD47 in apoptosis

CD47 is found to play an important role in the apoptotic process. Apoptosis is the programmed cell death which is the result of abnormal metabolism in the cell which creates the conditions where the survival of the cell becomes difficult or dangerous. Apoptosis is non-stopable once it is initiated. It is actually the result of the balance between the excitatory and inhibitory signal which are transmitted by these factors. It is very important for various types of activities like embryologic development, homeostasis maintenance. According to one hypothesis, the CD47 expression in the apoptotic cell is reduced due to which the macrophages attack them and digest them. Another hypothesis focuses on the fact that when CD47 and TSP-1 bind with each other, they make a new binding site for SIRPα. During apoptosis when this binding takes place, the phagocytic signal is released resulting in the phagocytosis of damaged cells. CD47 when bound to SIRPα or calreticulin releases an inhibitory signal which results in the suppression of the excitatory signal due to which the apoptosis of an abnormal cell does not initiate and it keeps on growing. If the CD47 is inhibited, the excitatory signal level goes high and it results in apoptosis induction. It
has been found that by inhibiting this CD47-SIRPα interaction, the tumor growth can be inhibited.  

**Role of CD47 in organ transplantation**

Apart from apoptosis it is found to be involved in organ transplantation. End stage organ failure can be cured by organ transplantation only. But due to lack of organ donors it was suggested that pig can act as a donor. But both innate as well as acquired immunity are responsible in xenograft rejection. Apart from immune cells, macrophages also play important role in this process. When murine CD47 was expressed in the porcine cells, the phagocytosis of these porcine cells by porcine macrophages was found to be decreased in vitro and the survival of the cell was found to be increased in vivo. Similar scenario was found when human macrophage SIRPα and porcine CD47 interactions were studied. Although being compatible with each other the porcine CD47 was found to be unable to activate human macrophage SIRPα due to which the porcine cells are rapidly phagocytosed by the human macrophages. So, it can be an excellent approach to produce xenogenic pigs expressing human CD47 and further minimize the xenograft rejection.

**Role of CD47 in Sickle cell anemia**

It is a genetic disease which is the result of a point mutation. The hemoglobin β chain is mutated due to which the formation of insoluble intracellular aggregates of hemoglobin takes place resulting in sickle shaped RBCs which is responsible for the vaso-occlusive crisis. The crisis is due to adhesion of the sickle erythrocytes to vascular endothelium. CD47 is found to be involved in the mechanism by binding to TSP.

**Role of CD47 in bone resorption**

Resorption is a process which is required for regeneration of adult skeleton. It is a continuous process where old bone tissue is continuously resorbed by osteoclasts which results in formation of bone cavity. That cavity is filled by the osteoblasts to form new bone matrix. This whole process is called bone remodeling which is very strictly controlled.

Koskinen *et al.* studied CD47−/− bone marrow culture and found that the production of M-CSF (Macrophage - Colony Stimulating Factor) and RANKL (Receptor Activator of Nuclear factor κβ Ligand) was reduced. M-CSF is required for the survival and proliferation while RANKL is responsible for the formation of the osteoclasts. Hence whole remodeling process is affected leading to disturbance of bone homeostasis. Osteoclast genesis was also found to be reduced in mutant SIRPα stromal cells. It clearly indicates that a CD47 induced SIRPα signaling is required for normal bone homeostasis.

**Biological functions of CD47**

Apart from its role in apoptosis CD47 also plays important role in some other biological processes. Some of them are enlisted below.

- In B lymphocytes, CD47 was suggested to synergistically promote cell migration
- CD47 interacts with SIRPα and inhibits the phagocytosis of tumor cells
- It interacts with thrombospondin-1 and inhibits nitric oxide signaling hence inhibits the angiogenesis and increases the risks of cardiovascular diseases eg. hypertension
- SIRPα and CD47 have been recently shown to be highly expressed in pancreatic β-cells. SIRPα mutant mice manifest a low plasma insulin level and impaired glucose tolerance, indicating that SIRPα promotes insulin secretion from β-cells. It remains unknown whether CD47 also participates in the effects of SIRPα in pancreatic β-cells or muscle, although expression pattern of CD47 hints that it is likely
- Co-ligation of CD47 and the T-cell antigen receptor (TCR) is a synergistic signal for activation of T-lymphocytes that requires CD47-induced association of protein kinase C with cytoskeleton
- CD47-SIRPα signaling pathway might participate in the regulation of hippocampus-dependent memory formation, of central
responses to stress and of the autonomic nervous system.\textsuperscript{14}

- Ligation of SIRP\alpha by CD47-Fc fusion proteins suppresses the phenotypic and functional maturation of immature dendrite cells (DCs) and inhibits the cytokine production.\textsuperscript{13}

**MATERIALS AND METHODS**

**Protein-protein interactions analysis**

Protein-protein interaction plays key role in predicting the protein function of target protein and drug ability of molecules. A variety of *in silico* methods have been developed to support the interactions that have been detected by experimental approach. While there have previously been some tools to model a few facets of biological systems, Schrödinger’s BioLuminate is the first comprehensive user interface and the lynchpin product of the Biologics Suite that is designed from the ground up, with significant user input, to specifically address the key questions associated with the molecular design of biologics. BioLuminate offers a state of the art protein-protein docking program, with modes for antibody and multimer docking.\textsuperscript{19}

**Molecular docking**

Molecular docking is a study of how two or more molecular structures, for example drug and enzyme or receptor fit together. The search algorithm should create an optimum number of configurations that include the experimentally determined binding modes. These configurations are evaluated using scoring functions to find the best binding configurations.\textsuperscript{20}

**Docking algorithms**

- *Genetic algorithms*

  Genetic algorithms and evolutionary programming are quite suitable for solving docking problems because of their usefulness in solving complex optimization problems. Some programs using genetic algorithms are GOLD, Auto Dock.

- *Incremental construction algorithm*

  The method involves dividing the ligand into fragments and docking them into active site, finally these fragments are linked together i.e. based on incremental construction algorithm. Selection of base fragment has been automated in newer programs such as FlexX and DOCK.\textsuperscript{21}

**Scoring functions for docking**

When the docking is completed the scoring function is used to rank each ligand in the database for which a docking solution has been found. The energy of binding is given by the Gibbs-Helmholtz equation:

\[
\Delta G = \Delta H - T\Delta S
\]

With $\Delta G$ giving the free energy of binding, $\Delta H$ the enthalpy, $T$ the temperature in Kelvin and $\Delta S$ the entropy. Bohn function is the type of scoring function which is most commonly used in docking software. It is based on calculation of $\Delta G$ given by following equation:

\[
\Delta G = \Delta G_0 + \Delta G_{\text{hb}} \sum f(\Delta R) f(\Delta \alpha) + \Delta G_{\text{lip}o, \text{A}} (\text{lipo}) + \Delta G_{\text{ion}} \sum f(\Delta R) f(\Delta \alpha) + \Delta G_{\text{rot N} \text{R}} + \Delta G_{\text{aro/aro}}
\]

$\Delta G_{\text{hb}}$ = free energy change due to hydrogen bonding

$\Delta G_{\text{ion}}$ = free energy change due to ionic interaction

$\Delta G_{\text{lip}o, \text{A}}$ = free energy change due to hydrophobic interaction

$\Delta G_{\text{aro/aro}}$ = free energy change due to aromatic-aromatic interaction

Glide is one of the widely used docking programs. It uses a series of hierarchical filters to search for possible locations in the active site region of the receptor. The properties of a receptor/active site region are represented by a grid that has different set of fields that provide progressively more accurate scoring of the
ligand pose. It uses a Glide Score (gscore) for predicting binding affinity and rank ordering of ligands in database screening. Glide approximates a complete systematic search of the conformational, orientation, and positional space of the docked ligand. Comparisons to published data on RMS deviations show that Glide is nearly twice as accurate as GOLD as and more than twice as accurate as FlexX for ligands having up to 20 rotatable bonds.22

**Molecular Dynamics**

It is computer simulation method based on Newton’s equations of motion for a system of interacting particles. Although normally represented as static structures, molecules are in fact dynamic. Most experimental properties, for example, measure a time average or an ensemble average over the range of possible configurations the molecule can adopt. One way to investigate the range of accessible configurations is to simulate the motions or dynamics of a molecule numerically. This can be done by computing a trajectory, a series of molecular configurations as a function of time, by the simultaneous integration of Newton's equations of motion:

\[
\frac{d}{dt} r_i(t) = v_i(t) \\
\frac{d}{dt} v_i(t) = F_i(t) \\
\frac{d}{dt} m_i
\]

for all atoms \((i = 1, 2, ..., N)\) of the molecular system. The atomic coordinates, \(r\), and the velocity, \(v\), of atom, \(i\), with mass, \(m_i\), thus become functions of time. The force \(F_i\) exerted on atom \(i\) by the other atoms in the system is given by the negative gradient of the potential energy function \(V\) which in turn depends on the coordinates of all \(N\) atoms in the system:

\[
F_i(t) = -\delta V(r_1(t)r_2(t)...r_N(t))
\]

(eqn. 3)

For small time steps \(\delta t\), eqn. (2) can be approximated by

\[
V_i(t+\Delta t/2) = v_i(t-\Delta t/2) + F_i(t) . \Delta t
\]

(eqn. 4)

and eqn. (1) likewise by

\[
ri(t+\Delta t) = ri(t) + vi(t+\Delta t/2)\Delta t
\]

(eqn. 5)

Equation (4) and (5) form the so-called leap-frog scheme for integrating Newton's equations of motion. Typically a time step of 1 to 10 fs is used for molecular systems. Thus a 100 ps (10^{-10} seconds) molecular dynamics simulation involves 10^5 to 10^4 integration steps. Even using the fastest computers only very rapid molecular processes can be simulated at an atomic level. As with any aspect of modeling, the accuracy of the predicted dynamics will depend on the validity of the underlying assumptions of the model. In this case the model is essentially defined by the force field that is used.

In order to reproduce the actual behavior of real molecules in motion, the energy terms described above are parameterized to fit quantum-mechanical calculations and experimental (for example, spectroscopic) data. This parameterization includes identifying the ideal
stiffness and lengths of the springs that describe chemical bonding and atomic angles, determining the best partial atomic charges used for calculating electrostatic-interaction energies, identifying the proper van der Waals atomic radii and so on.

Collectively, these parameters are called a ‘force field’ because they describe the contributions of the various atomic forces that govern molecular dynamics. Several force fields are commonly used in molecular dynamics simulations, including AMBER, CHARMM, and GROMOS. These differ principally in the way they are parameterized but generally give similar results. Once the forces acting on each of the system atoms have been calculated, the positions of these atoms are moved according to Newton’s laws of motion. The simulation time is then advanced, often by only 1 or 2 quadrillionths of a second, and the process is repeated, typically millions of times.²³

Factors that govern the outcome of MD simulations are:

i. choice of the degrees of freedom
ii. force field parameters
iii. treatment of non-bonded interactions
iv. solvation effects
v. boundary conditions
vi. treatment of temperature and pressure
vii. integration time step
viii. starting configuration

Desmond is a software package developed by D.E. Shaw Research to perform high speed molecular dynamics simulations of biological systems. The code used novel parallel algorithm and numerical methods to achieve high performance on multiprocessor systems. Long-range electrostatic energy and forces are calculated using particle mesh Ewald-based methods. Constraints are enforced using the M-SHAKE algorithm. These methods can be used together with time-scale splitting (RESPA-based) integration schemes.²⁴

**Foldamers**

A foldamer is a chain molecule or oligomer that folds into a conformationally ordered state in solution. They are the molecules that mimic the ability of proteins, nucleic acids, and polysaccharides to fold into well-defined conformations, such as helices and β-sheets. The structure of a foldamer is stabilized by noncovalent interactions between non-adjacent monomers.

**Approaches toward foldamers designing**

There are many approaches which are available for foldamers designing. Among them homologation approach and sequence based designing approach are the most widely used till now.

**Homologation approach:** In this approach either one or two extra -CH₂ groups are added in the peptide backbone while the side chains of the amino acids are kept intact. The nomenclature of the designed foldamers vary according to the position of the side chain of the foldamer attached to it. β² foldamers are the foldamers which are having their side chain attached to the -CH₂ which is present just next to Cα, as shown in figure below. Similarly, β³ peptides are designed by attaching the side chain to the -CH₂ prior to -NH moiety. γ peptides are designed by inserting two -CH₂ groups in the backbone and the side chain is attached to the -CH₂ which is in vicinity of -NH moiety.

---

![α-peptide](image-url)
Sequence based designing: In this approach the backbone of the peptide is partially modified resulting into the resistance against the protease attack. Rather than increasing the backbone length of the peptide by inserting -CH2 groups in it, sequence based designing is more focused on inserting unnatural moieties in the backbone.18

A total of 47 protein-foldamer complex structures are available in PDB as of 26 May 2018. The list of the structures is given below. Most of these structured are determined by X-Ray diffraction method except one structure (PDB ID: 5US3) and all are targeted against the protein while only one structure is targeted for DNA G-Quadruplex (PDB ID: 5HIX). No structure was found where CD47 was targeted with foldamers. They are based on homologation as well as sequence based approach. Most of the structures are β2 and β3 based and no structure was found for γ4 foldamer in PDB. Most of the foldamers in these structures are containing at least 13 units. No structure is available in PDB which contains small foldamers.

RESULTS AND DISCUSSION

Comparison of bound and unbound structures of CD47 and SIRPα
We selected PDB structure 2JJS (1.85Å) for analysis. It contains the complex structure of CD47 and SIRPPα which is available in dimer form. The protein was prepared using Protein preparation wizard in Maestro. Hydrogens were added during the preparation. The extra chains were removed and only one chain for CD47 (chain c) was kept. Then the protein was energy minimized. Then this protein complex was superimposed on the chain C which is obtained from 2JJS without energy minimization. The RMS between both the structures was found to be 0.1727 Å for 113 Cα atoms. It means that the CD47 does not undergo any specific conformational change during the ligand binding.
**Protein-protein interaction analysis**
To determine the important residues which are binding with SIRPα we performed protein-protein interaction analysis. We used Bioluminate module of Schrodinger for this purpose.

**Role of Hydrogen bonds**
Hydrogen bond formation takes place when one hydrogen atom which is connected to an electronegative atom comes in contact with another electronegative atom. It is mainly of two types i.e. Intramolecular and intermolecular H-bonding.

**Table 1. List of H-bonds which are taking place between CD47 and SIRPα residues.**

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>CD47 residue</th>
<th>SIRPα residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>T102</td>
<td>L30</td>
</tr>
<tr>
<td>2.</td>
<td>L101</td>
<td>G34</td>
</tr>
<tr>
<td>3.</td>
<td>E104</td>
<td>Q52</td>
</tr>
<tr>
<td>4.</td>
<td>Pca1</td>
<td>S66</td>
</tr>
<tr>
<td>5.</td>
<td>R103</td>
<td>T67</td>
</tr>
<tr>
<td>6.</td>
<td>E29</td>
<td>R69</td>
</tr>
<tr>
<td>7.</td>
<td>T102</td>
<td>K93</td>
</tr>
<tr>
<td>8.</td>
<td>K39</td>
<td>G97</td>
</tr>
<tr>
<td>9.</td>
<td>D46</td>
<td>S98</td>
</tr>
</tbody>
</table>

Among these T102 is the most important CD47 residue because it is responsible for making two H-bonds with SIRPα. Some important H-Bond interactions are shown below.
Figure 5. Representation of H-Bonds between T102 (CD47), L30 (SIRPa and K93 (SIRPa).

Figure 6. Representation of H-Bonds between Pca1 (CD47) and S66 (SIRPa).

Figure 7. Representation of H-Bonds between L101 (CD47) and G34 (SIRPa).
Role of salt bridges
Salt bridge is the combination of two types of bonding i.e. Hydrogen bonding and ionic bonding. The CD47 residue E97 is the most important salt bridge forming residue which interacts with the SIRPα residues. R69 is the main SIRPα residue which is found to form two salt bridges with E35 and E100 of CD47. We did not find any kind of van der Waals or π-π interaction in the complex. It is found that SIRPα is binding to CD47 at different residues which are not present in a sequence. The residues which are present in the loops are actively involved in binding with SIRPα. T102 is the outermost residue of CD47 and it is found to make two Hydrogen bonds with L30 and K93 of SIRPα. The loop containing L101, T102 and R103 is the most active one. This single loop alone is responsible for more than 6 interactions with SIRPα residues.

Table 2. List of CD47 and SIRPα residues undergoing salt bridge formation.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>CD47 residue</th>
<th>SIRPα residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>E104</td>
<td>N51</td>
</tr>
<tr>
<td>2.</td>
<td>E97</td>
<td>K53</td>
</tr>
<tr>
<td>3.</td>
<td>E97</td>
<td>K96</td>
</tr>
<tr>
<td>4.</td>
<td>K6</td>
<td>E54</td>
</tr>
<tr>
<td>5.</td>
<td>E104</td>
<td>Q52</td>
</tr>
<tr>
<td>6.</td>
<td>E35</td>
<td>R69</td>
</tr>
<tr>
<td>7.</td>
<td>E100</td>
<td>R69</td>
</tr>
</tbody>
</table>

Figure 8. Representation of salt bridges between K6 (CD47) and E54 (SIRPα).

As a whole it can be said that residues Pca1, Glu35, Lys39, Asp46, Glu97, Thr99, Glu100, Leu101, Thr102, Arg103, Glu104 and Glu106 are the most important residues which are binding with SIRPα.
Figure 9. Representation of salt bridges between E97 (CD47) and K96 (SIRPα).

Figure 10. Representation of the active residues of CD47.

**SiteMap analysis of CD47**

We took PDB ID: 2JJS for SiteMap Analysis. The protein was prepared by using maestro module of Schrodinger software. It contained four chains labelled as A, B, C and D. Out of them C and D chains represented the CD47 while A and B chains represented SIRPα. C and D chains were homologous in nature. We removed all the chains except C. Then we removed N-acetyl glucosamine and Iodide ion and generated state. After that we optimized H-Bonding and sample water orientation. Then the protein was energy minimized and SiteMap analysis was performed in which we found that the SiteScore for the CD47 is 0.494 while Dscore is 0.423 which is very less for designing any small molecule against it.
**Foldamer designing**

A total of 270 foldamers were designed using homologation and sequence based approaches. Because of lack to the availability of any the peptide binding to CD47 we had to consider the important binding residues of the CD47 and SIRPα. Then on the basis of the amino acid residues present in SIRPα which are very crucial in binding with the CD47, the different foldamers were prepared. We basically considered the residues which were involved in strong binding like Hydrogen bonding and salt bridge formation like Pca1, Glu35, Glu97, Thr99, Glu100, Thr102, R103, Glu104 and Glu106. Foldamers were drawn using Chemdraw 12.0. Then we minimized the drawn foldamers using ChemBio3D using MMFF94 force field. Then the molecules were saved in .sdf format.

**Molecular docking studies**

**Protein preparation**

2JJS was downloaded from PDB. It contained four chains A, B, C and D. Hydrogen atoms were added, water molecules were removed and missing residues were repaired using prime. Then chains A, B, and D were removed. Then the protein was H-bond optimised and energy minimized using OPLS-2005.

**Grid generation and docking**

Grid was generated using the following residues: Pca1, Glu35, Glu97, Thr99, Glu100, L101 Thr102, R103, Glu104 and Glu106. These are the most important residues which are taking place in the binding with the CD47. The centroid of the selected residues was kept as the centre of the grid. Since our molecules were containing more than 150 atoms and the number of rotatable bonds were also larger than normal molecule. So, considering these factors, the size of the grid box was kept as (20Å x 20Å x 20Å). Then docking was performed using Glide. We selected SP-peptide precision mode. Then docking was performed for all the designed molecules. The top 10 docking results are enlisted in Table no. 4.

**Table 3. Scores of top 10 designed foldamers.**

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Type of foldamer</th>
<th>Docking score (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>γ⁴</td>
<td>-8.325</td>
</tr>
<tr>
<td>2</td>
<td>γ⁴</td>
<td>-8.081</td>
</tr>
<tr>
<td>3</td>
<td>γ⁴</td>
<td>-8.001</td>
</tr>
<tr>
<td>4</td>
<td>γ⁴</td>
<td>-7.779</td>
</tr>
<tr>
<td>5</td>
<td>β²</td>
<td>-7.670</td>
</tr>
</tbody>
</table>
We had designed foldamers with as low as 4 amino acid but due to lack of good docking score we had to increase the length of the foldamers. At last we were able to get the docking score above -8.000 only with the foldamers which were having backbone equal to 6 amino acids. Out of top 10 results, first 4 foldamers are $\gamma^4$ derivatives only and all of them are designed by sequenced based approach. No foldamer was found in top results which is based on homologation approach. The structures of top 10 foldamers with best docking score is given on the next page.
**2D interaction diagrams**

2D interaction diagram gives us an idea about the residues which are involved in the binding of our ligand. The type of interaction is also shown by different colors e.g. magenta lines used in images below show presence of H-bonds between ligand and amino acid residue.
In Figure 12 (A) we can see that Pca1, Glu35, Thr99, L101, Thr102, Glu104 and Glu106 are making direct H-bond with the ligand. Similarly in Figure 12 (C), the ligand is making H-bonds with Glu97, Leu101, Thr102, Glu104 and Glu106 of CD47 but Foldamer 2 is not making that much good interactions with the protein. Since this foldamer contains proline which is a cyclic amino acid as the end terminal, the flexibility of the side chain is lost. So it may be a reason for lesser number of significant contacts.

**Molecular dynamics studies**

Three molecules were selected for the comparative molecular dynamic study. These molecules were selected on the basis of Glide Score of the respective molecule along with number of interactions between ligand and receptor.

**System building:** The solvent model was kept TIP3P and orthorhombic box was selected with buffer as box size calculation method. The system was neutralized by adding 2 sodium ions (Na+). The salt concentration was set to 0.15M using Na+ and Cl- ions.

**Molecular Dynamics:** All simulations were performed up to 15ns with 15ps recording interval at constant temperature (300 K) and pressure (1.01325 atm). Periodic boundary conditions were applied and surface tension was kept 4000Å. Bennett acceptance ratio (BAR) method was used to estimate the free energy difference from the MD simulations. Similarly, the whole process was repeated for another two molecules. Then the simulations of the docked complex structures of CD47 with foldamer 1, foldamer 2 and foldamer 3 were performed to obtain the dynamical pictures of the conformational changes that occur in aqueous solutions. The RMSDs of the trajectory of protein and ligand are depicted below.

**Protein RMSD**

The Root Mean Square Deviation (RMSD) is used to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. It is calculated for all frames in the trajectory. The plots below in Figure 13 show the RMSD evolution of a protein (left Y-axis). All protein frames are first aligned on the reference frame backbone, and then the RMSD is
calculated based on the atom selection. Monitoring the RMSD of the protein can give insights into its structural conformation throughout the simulation. RMSD analysis can indicate if the simulation has equilibrated - its fluctuations towards the end of the simulation are around some thermal average structure. Changes of the order of 1-3Å are perfectly acceptable for small, globular proteins. Changes much larger than that, however, indicate that the protein is undergoing a large conformational change during the simulation.

In all the three simulations the RMSD of the protein is about 1.2Å that means the protein is not undergoing large conformational changes. After 12ns the protein starts stabilizing and the change in RMSD minimizes at this time. Although there are slight deviations in Foldamer 3 complex (Figure 13 (C)).

**Ligand RMSD**

Ligand RMSD (right Y-axis) indicates how stable the ligand is with respect to the protein and its binding pocket. In the above plot, 'Lig fit Prot' shows the RMSD of a ligand when the protein-ligand complex is first aligned on the protein backbone of the reference and then the RMSD of the ligand heavy atoms is measured. If the values observed are significantly larger than the RMSD of the protein, then it is likely that the ligand has diffused away from its initial binding site. In Figure 13 (A) it can be seen that RMSD of ligand is near about 7Å after the stabilization of the complex i.e. after 12ns. Similarly, in Figure 13 (B) and (C) the RMSDs are 7.5Å and 9Å respectively. This much large deviation in RMS is due to the presence of a number of rotatable bonds in our ligands.
Figure 13. (A) Plot of RMSD of docked complex of CD47 and foldamer 1 versus MD simulation time in MD simulated structure. (B) Plot of RMSD of docked complex of CD47 and foldamer 2 versus MD simulation time in MD simulated structure. (C) Plot of RMSD of docked complex of CD47 and foldamer 3 versus MD simulation time in MD simulated structure.

Protein RMSF
The Root Mean Square Fluctuation (RMSF) is useful for characterizing local changes along the protein chain. In this plot, peaks indicate areas of the protein that fluctuate the most during the simulation. It is observed that the tails (N and C-terminal) fluctuate more than any other part of the protein. Secondary structure elements like alpha helices and beta strands are usually more rigid than the unstructured part of the protein, and thus fluctuate less than the loop regions.

Figure 14. (A) RMSF of Protein when bound to foldamer 1. (B) RMSF of Protein when bound to foldamer 2. (C) RMSF of Protein when bound to foldamer 3.
In Figure 14 (A), the residues present between residue index 20 and 40 show high fluctuations in the protein RMSF. It shows the presence of the loop region in that range.

**Protein-ligand contacts**

Protein interactions with the ligand can be monitored throughout the simulation. These interactions are categorized by type and summarized as shown in the plot above. Protein-ligand interactions (or contacts) are categorized into four types: hydrogen bonds, hydrophobic (π-cation and π-π interaction), ionic and water bridges.
The value 0.5 indicates 50% of the simulation time the specific interaction was maintained. In figure 15 (A) it can be seen that the Pca1, Glu35, Thr103 and Glu104 are making the interactions which were maintained for more than 50% of the simulation time.

**Figure 15.** (A) Protein-ligand contacts between CD47 and foldamer 1 during simulation. (B) Protein-ligand contacts between CD47 and foldamer 2 during simulation. (C) Protein-ligand contacts between CD47 and foldamer 3 during simulation.
Figure 16. (A) Stability of protein-ligand contacts between CD47 and foldamer 1. (B) Stability of protein-ligand contacts between CD47 and foldamer 2. (C) Stability of protein-ligand contacts between CD47 and foldamer 3.

In Figure 16 (A), (B) and (C) the top panels show the total number of specific contacts the protein makes with the ligand over the course of the trajectory. The bottom panel shows which residues interact with the ligand in each trajectory frame. Some residues make more than one specific contact with the ligand, which is represented by a darker shade of orange, according to the scale to the right of the plot.

In Figure 16 (A) it can be seen that Pca1 and Thr102 are making contacts with the ligand almost throughout the simulation. Glu35 is also making contacts during more than half the simulation time. In Figure 16 (B), it can be seen that Glu34 and Glu37 are in contact with the ligand rather than Glu35 as in previous case. Here Asp51 is making strong interaction with the ligand. In Figure 16 (C), Glu97 and Glu106 are the main residues which are making the contacts with the ligand. Thr99 and Glu104 also make contacts but they are not as continuous as the contacts with Glu97 and Glu106.

**CONCLUSIONS**

Development of the effective anti-cancer molecules has been a challenging task always. In this study we have tried to target CD47, one of the least explored target in the field of cancer. During SiteMap analysis we found that CD47 contains no active site for which we can design small molecules while antibodies have their own problems. So, rather than small molecules or antibodies we have taken foldameric approach for this purpose.

We have studied protein-protein interaction between CD47 and SIRPα. Then we selected the important residues which are playing crucial role in CD47 binding. Most of these residues are either...
forming H-bond or salt bridge with SIRPα. Then we designed 270 different types of foldamers on the basis of those selected residues. The foldamers were designed based upon both the approaches i.e. homologation as well as sequence based approach. In this study these foldamers were docked upon CD47. The best docking score was found to be -8.325. In the last objective molecular dynamics of three designed molecules with top docking scores were performed for 15ns to understand the stability of the interactions. The interactions were found to be stable in two out of three simulated CD47-foldamer complexes.

**ABBREVIATIONS**

PDB, Protein Data Bank; RMSD, Root Mean Square Deviation; CD, Cluster of Differentiation; SIRP, Signal Regulatory Protein; RMSF, Root Mean Square Fluctuation; MD, Molecular Dynamics; Pca, Pyroglutamic acid; Ig, Immunoglobulin; IAP, Integrin Associated Protein; MCSF, Macrophage - Colony Stimulating Factor; TSP, Thrombospondin; ps, Picosecond; fs, femtosecond; ATM, Atmosphere; BAR, Bennett acceptance ratio; PSA, Polar Surface Area; SASA, Solvent Accessible Surface Area; rGyr, Radius of Gyration.

**REFERENCES**


How to cite this article:

Source of Support: Nil

Conflict of Interest: None declared

Your next submission with British BioMedicine Publishers will reach you the below assets
- Quality Editorial service
- Swift Peer Review
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  - (Pdf, E-pub, Full Text)
- Unceasing customer service

Track the below URL for one-step submission
http://www.britishbiomedicine.com/manuscript-submission.aspx