

ABSTRACT

We have designed de novo a simple, linear model of Vitamin-D binding protein derived macrophage activating factor (GcMAF), using software Insight II. Domain III of chain A in vitamin D was isolated and bonded with oligosaccharide moiety which is composed primarily of N-acetyl-D-galactosamine (GalNAc), galactose and sialic acid residues. Activation of Gc-MAF is accomplished by selective removal of two sugars galactose and sialic acid. Recent discoveries have shown that small peptide domains will allow the introduction of peptides and proteins into cells so, the structure was altered to contain only twenty-one amino acids with one sugar N-acetylgalactosamine attached to Threonine. Then a linear polypeptide was modeled which resembles the structure of small peptide domain created. Both the structures were minimized to be as small as possible retaining all the features of the structural element and then compared. This simple model system will help in creating synthetic peptide domains, which can be introduced into cells.

INTRODUCTION

Vitamin-D binding protein derived macrophage-activating factor (GcMAF)

Vitamin D binding protein (DBP), also known as a group specific component (Gc), is a polymorphic serum glycoprotein with multiple functions. Vitamin D binding protein (DBP) is a 52Kda protein that binds monomeric actin in addition to vitamin D. DBP has been detected on the surface of several cell types, yolk sac endodermal cells, and some T-lymphocytes. In B-cells, DBP seems to participate in the linkage of surface immunoglobulins.

The protein is 458 residues in length (Cooke, 1986), and forms three domains, the first of which contains the sterol binding site. The three domains share limited sequence homology, with each other and to similar repeats in human serum albumin. The affinity for actin monomers is high and the actin binding site has been reported to reside within domain III between residues 350 and 403 (Haddad *et al*, 1992). The structure of the complex of DBP and actin (Otterbein *et al*, 2002) confirms that domain III forms an actin-binding contact, between sub-domains 1 and 3 of actin.

Gc-MAF is an evolutionarily conserved serum glycoprotein composed of three distinct domains Domain III is O-glycosylated and the oligosaccharide moiety is composed primarily of N-acetyl-D-galactosamine (GalNAc), galactose and sialic acid residues (Viau *et al*., 1983). Activation of Gc-MAF is accomplished by selective removal of sugars by galactosidase and sialidase by activated B- and T-cells, respectively (Fig. 1). Retention of a single GalNAc residue is essential for activity. The product Gc-MAF activates macrophages through an interaction of the GalNAc residue with a C-type lectin receptor on the macrophage surface (Iida *et al*., 1999).

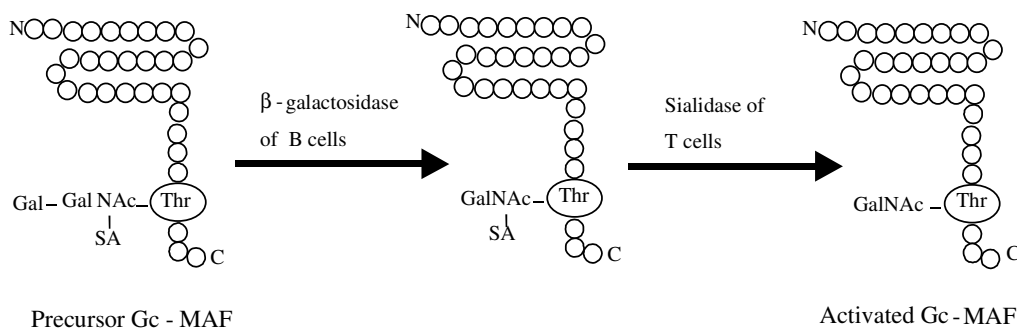


Figure 1. In vivo activation of Gc-MAF by selective removal of sialic acid and galactose. (For in vitro activation, immobilized galactosidase and sialidase will be used)

Cancerous cells secrete α -N-acetyl-D-galactosaminidase (GalNAcase) into the blood stream, which results in complete deglycosylation of serum Gc-MAF leading to immunosuppression (Koga *et al.*, 1999; Yamamoto and Naraparaju, 1997). It has been shown that the administration of enzymatically activated Gc-MAF to mice will overcome the inactivation and result in macrophage activation in less than 6 hr. Injection of Gc-MAF also substantially increases initiation of antibody production within 48 hr. These observations show that Gc-MAF can be useful as an adjuvant to enhance and accelerate the development of the immune response and to generate a large amount of antigen-specific antibodies. Although activated MAF results in greatly enhanced macrophage phagocytic activity and consequently has utility as a therapeutic agent, particularly in individuals afflicted with cancer, its role may be more important in T-cell stimulation and antibody production. The only current source of Gc-MAF is mammalian serum, which restricts its usage.

Gene therapy has been advocated as an ideal approach to introduce molecules into cells. This approach involves transfecting cells with a gene that produces a specified protein. However, in many cases, the effectiveness of gene therapy has been limited by immunogenicity of the viral vectors, targeting the vectors to the appropriate cells, and regulation of expression of the gene so that the correct amount of protein is produced and targeted in the cell. Finally, with gene therapy there is a delay in the expression of the target protein. Many of these problems can be obviated by directly transducing the target protein into the cell .

The major limitation of protein-based therapies has been introducing the protein into the cells. However, recent discoveries have shown that small peptide domains will allow the introduction of peptides and proteins into cells. So, we are using molecular modeling to create such small polypeptide of GcMAF protein ,which can be introduced into cells.

Cardiovascular and cancer are the leading causes of death in US and in the state of Arizona. Recently, two novel therapeutic proteins with great promise of limiting mortality caused by these diseases, have been discovered. Until now, mammalian systems have been the only source of these two and most other therapeutic proteins. However, there are concerns of possible disease transmission from contaminating viruses, prions and other infectious agents when these therapeutic proteins are produced in mammalian systems. Plants are able to produce human proteins that have therapeutic applications and do not pose the risk of transmitting animal diseases. Plants can be grown at large-scale to produce kilogram quantities of medically important proteins for patient care. We will primarily use tobacco plants as 'green biofactories' develop recombinant protein therapeutic modalities. This represents a novel approach that has significant biotechnology potential.

MOLECULAR MODELING:

Molecular modeling is a collection of computer based techniques for deriving, representing and manipulating the structures and reactions of molecules, and those properties that are dependent on these three dimensional structures. In a natural way molecular modeling treats a molecule as a collection of weights connected with springs, where the weights represent the nuclei and the springs represent the bonds.

Using professional programs and supercomputers one can search for the most stable structure of a given molecule (so-called **energy minimization**) or simulate its oscillations at any given temperature (**molecular dynamics**). These techniques are useful for the refinement of conformations of molecules determined by physical methods. It is far more challenging to employ these methods to predict the conformation, dynamics and properties of molecules whose structure is not yet determined by independent methods.

Molecular modeling provides a possibility of screening for those modifications, which are most likely to be successful. Molecular modeling is very useful in predicting the alignment of molecules in complexes. It's important in designing new drugs, which have to interact in a desired way with, for instance, cell receptors.

SOFTWARE PACKAGES:

These are some examples of specialized software packages used for molecular modeling.

- *Iditis*, *PROTEP*, *PDBtool* and *ChemDBS-3D* are used for structure database manipulation.
- *O*, *XtalView*, *MOIGEN+*, *CONCORD* and *Cobra* are used for model building.
- *RasMol*, *ChemDraw*, *ISIS Draw*, *Raster3D*, *VMD* and *Kekule* are used for Structure drawing and visualization.
- *CHARMm*, *Discover* and *Polaris* are used for molecular mechanics.
- *MacroModel*, *QUANTA/CHARMm*, *SYBYL*, *Cobra* and ***InsightII*** are used for molecular dynamics.
- *DGEOM*, *X-PLOR* and *SYBYL* are used for Distance geometry.
- *GAMESS*, *HONDO* and *UniChem* are used for Quantum chemistry.
- *NMRchitect*, *X-PLOR* and *DGEOM* are used for NMR-based structure determination.

In our project, we have used the ***Insight II*** program, which is part of the MSI software. *Insight II* is a 3D graphical environment for molecular modeling. *Insight II* environment integrates builder modules, development tools, simulation and visualization tools with tools specifically developed for applications in the life and materials sciences.

INSIGHT II AND MODULES:

Insight II runs on Silicon Graphics workstations and servers. Insight II creates, modifies, manipulates, displays, and analyzes molecular systems and related data and provides the basic requirements for all Insight II software modules. There are several modules some of which are *Viewer*, *Builder/Biopolymer*, *Ampac/Mopac*, *Discover*, *Discover_3*, *Analysis*, and *Docking*.

The modules we used in modeling GcMAF are:

Biopolymer constructs models of peptides, proteins, carbohydrates, and nucleic acids for visualizing complex macromolecular structures and for use in further simulation work.

Discover is a simulation program available within Insight II. It include dynamics simulations, minimization, and conformational searches, allow to predict the structure and properties of organic, inorganic and biological systems.

Builder constructs models of inorganic and organic crystals, metals, layers, and interfaces, which are used to develop and validate models of solids and surfaces.

METHODOLOGY

MODELING Gc-MAF USING INSIGHT II:

The required crystal structure of the complex of actin with Vitamin D- Binding Protein along with its CHAINS, Chain-A and Chain-B was first obtained from NCBI-Structure.

The Vitamin D protein has essentially two chains, Chain A and Chain B. Our study of interest was *Chain A- Domain III* as it consisted of amino acid residue *threonine* at position A418 which carried a *trisaccharide* composed of *N-acetylgalactosamine*, *galactose* and *sialic acid*. After separating domain III, the trisaccharide was created and attached to the free end of threonine using software Insight II. A series of Energy minimizations and dynamics were performed on this structure until it reached a global minimum of zero.

The structure was further analyzed by chopping amino acids and removing two sugars *galactose*, *sialic acid* such that the structure was as small as possible while retaining all its features, with minimum number of interactions.

In the next step, a polypeptide was created exactly with the same number of amino acids and trisaccharide was attached to the free end of the threonine in the same manner as above. This was mainly to verify whether it had the same structure as the original one so that a synthetic polypeptide can be created which would be directly attached to the activation site.

PROCEDURE:

The following are the **steps** used in Insight II for modeling the above-discussed protein (GcMAF):

1) Getting the PDB data file into the INSIGHT II environment:

The required crystal structure of Vitamin D- Binding Protein along with its chains, Chain-A and Chain-B was first obtained from NCBI (PDB – entry 1LOT).

2) Separating the Domain III molecule:

The following are the commands used in Insight II.

Start from the Main Menu:

- **Display the molecule.**

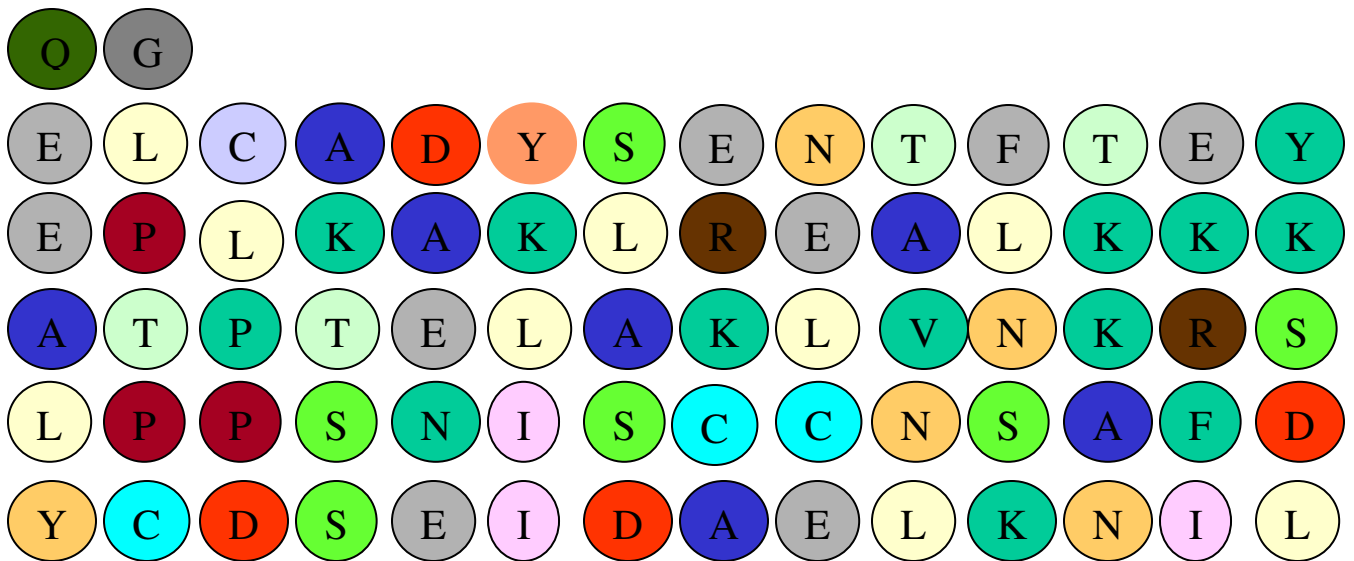
In order to get the molecule to the work frame we used the following command.

Command: MOLECULE → GET → Specify the required molecule.

- **Getting the sequence.**

The sequence of domain III amino acids was identified from the table.

The following was the sequence of CHAIN A – DOMAIN III identified in the table:



This was indicated in the table as A387-A457

Command: MOLECULE → TABULATE → Specify the above sequence.

- **Coloring the sequence**

We then colored this sequence in order to identify the Domain III from the protein.

Command: MOLECULE → COLOR → Specify the above sequence and select the required color.

Command: MOLECULE → DISPLAY → Specify the chains and domains other than Chain-A Domain III.

Once the different chains/domains were displayed *delete* * in the command prompt deleted the specified. In this way, all the domains other than the CHAIN A, DOMAIN III were removed from the structure. Then only the required domain III was left.

- **Saving the file**

A series of steps were involved in saving a file.

Command: OBJECT → RENAME → Rename the molecule.

- **Choose the BUILDER Module**

Setting Lone Pairs:

Command: MODIFY → HYDROGENS → Check the lone pairs.

Potentials Fixing:

From left Frame select FF Icon → Potentials → Fix Potentials.

- **Choose the DISCOVER Module**

Minimization:

Command: PARAMETERS → MINIMIZE → Set the default values and click execute.

Command: RUN → Run the Minimization.

Note: We made sure that the molecule name was listed and 'minimize' was selected, and then we executed the menu. The molecule must undergo some kind minimization in order to be saved. It usually takes the .car extension if it is saved properly.

3) Creating Sugar Molecules:

The Sugar molecules are:

- Sialic acid
- N-acetyl galactosamine (GalNAc)
- Galactose (Gal)

The Structure of the three sugars with bonds is as shown in Figure 2

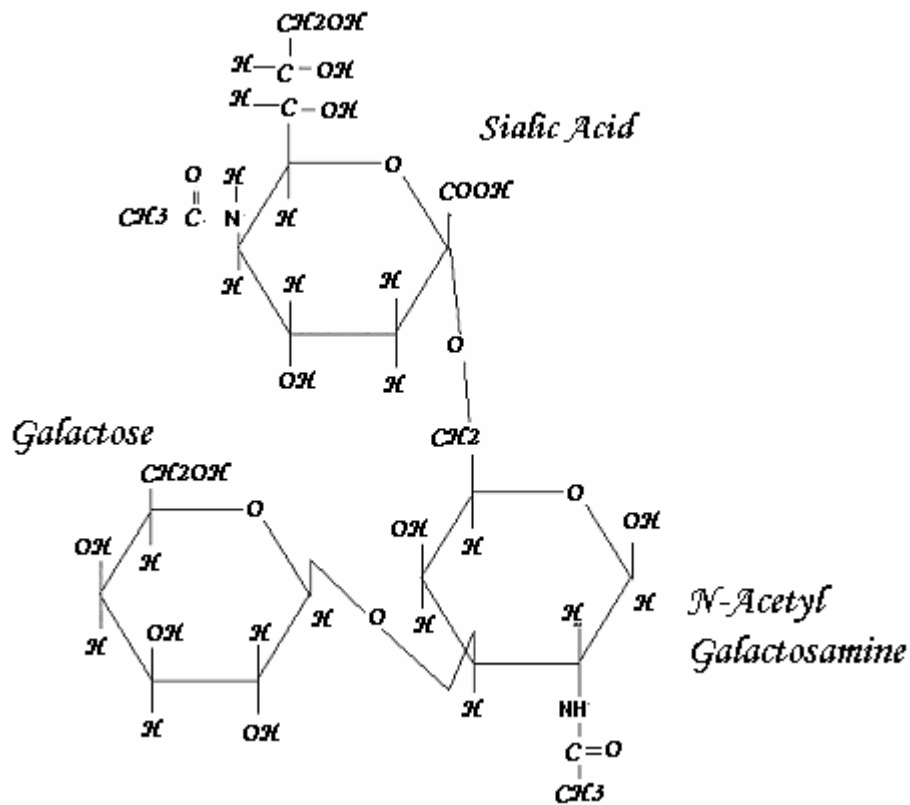


FIGURE 2: Sugar Molecule (Galactose(β 1-3), Sialic Acid (α 1,6), GalNAc(α 1-O)-Thr,)

The structure in Figure2 was built in INSIGHT II as follows:

- **Choose the BUILDER module**

Command: **TOOLS** → **MOL BUILDER** → **Select appropriate elements** → **-Bonds/Rings**

Command: **2D** → **3D** in **MOL BUILDER** turns the whole molecule into a 3D image.

- **Choose the DISCOVER Module.**

Command: **PARAMETERS** → **MINIMIZE** → **Set the default values and click execute.**

Command: **RUN** → **Run the Minimization.**

Minimization steps were repeated until global minimum was zero. This completed the building of the sugar molecule.

4) Bonding the Oligosaccharide moiety and the Domain III:

The bonding of the molecule was carried out by opening the desired molecules on the same window. The OH group of N-acetylgalactosamine(GalNAc) molecule at the beta end was removed. Then a bond was created between two hydrogen atoms of GalNAc and Threonine residue of DomainIII. The steps are as follows:

- **Choose BUILDER Module.**

Command: **BOND** → **Specify the Hydrogen atom from GalNAc and link it to the free end of the threonine of the DOMAIN III.**

Once the Bond gets created the sugar molecule attaches to the domain III molecule. Then the side chains of Domain III were fixed and the structure was minimized so that the sugar molecules could move around Domain III. This was done as follows:

- **Choose the DISCOVER Module.**

Command: **CONSTRAINTS** → **Specify the following**
Activation: Add
Atoms_To_Fix : Side chains
Constraint_Level: Molecule

- **Choose the DISCOVER Module.**

Command: **PARAMETERS** → **MINIMIZE** → **Set the default values and click execute.**

Command: **RUN** → **Run the Minimization.**

Repeat the minimization until the global minimum is zero.

The obtained structure after minimization was modified by

- **Removing two sugars** – Sialic Acid and Galactose
- **Making Domain III as small as possible** by just taking twenty-one amino acids i.e. ten on either side of *Threonine*. This structure was then minimized until a global minimum reached zero.

Next we modeled a linear polypeptide domain III of chain A and compared this to the minimized molecule in section 4. We again analyzed this structure by attaching the trisaccharide sugar molecule and next by removing two sugars from the trisaccharide.

5.) Modeling linear Polypeptide:

The polypeptide of Chain A-domain III was created as follows:

- **Choose BIO POLYMER Module.**

Command: **RESIDUE** → **APPEND** → **MOTIF (select the desired helix)**

The residues were added according to the actual amino acid sequence.

Adding lone pairs, minimizing the molecule and finally saving the file was carried out as discussed in above sections.

Command: **RESIDUE** → **CAP** → **N and C Terminus**

The created linear polypeptide was then capped with NH₂ in the N terminus and COOH in the C terminus.

This completed the building of linear polypeptide.

6) Attaching sugar to the polypeptide:

Attach the originally created oligosaccharide to the modeled polypeptide. The procedure for bonding the two molecules was same as discussed in the previous section.

The whole molecule was minimized for a few iterations until the global minimum reached zero. The dynamics of the molecule were also lowered besides minimizing. This was done as follows:

- **Choose DISCOVER module.**

Command: **PARAMETERS** → **DYNAMICS** → **Specify the following**
Equilibration : 100
Steps : 10000
Temperature : 300

Command: **RUN** → **Run_Dynamics**

After minimizing this structure galactose and sialic acid were removed from the oligosaccharide moiety. After running several minimizations and dynamics the structure slowly started to resemble the small peptide domain that was created earlier. We need to run some more dynamics before actually confirming the structure. Once that structure is obtained we can then hopefully design a synthetic polypeptide.

RESULTS:

The results obtained after performing the above discussed methods are:

- The Domain III of chain A after deleting remaining chains looks like the Figure below (Figure 3) .The blue colored part is the Threonine residue to which the oligosaccharide was attached. We can clearly see that the Threonine end is sticking out.

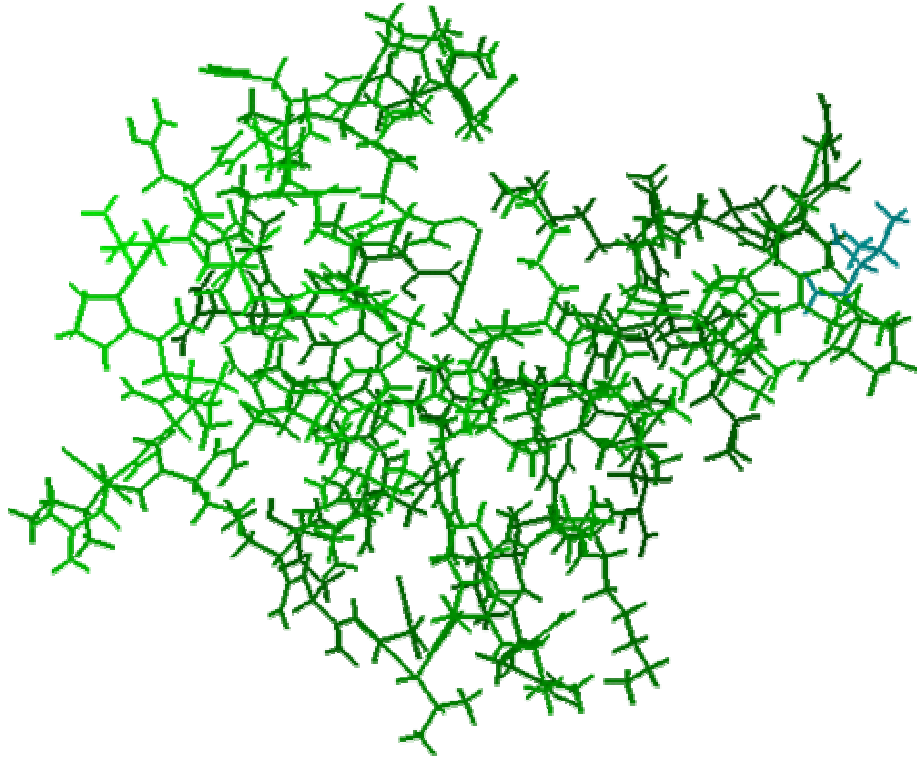


FIGURE 3: DomainIII

- The oligosaccharide created using INSIGHT II looks like the Figure below (Figure 4). The three sugars Galactose, GalNAc and Sialic acid can be seen as red, green and brown respectively.

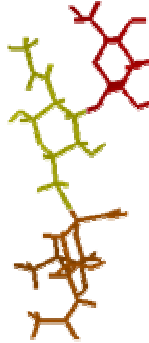


FIGURE 4: Oligosaccharide (Galactose(β 1-3), Sialic Acid (α 1,6), GalNAc)

- Then the sugar molecule which was created in INSIGHT II as shown in Figure 4 was attached to the Domain III Molecule (Figure 3). The result is as seen in Figure below (Figure 5)

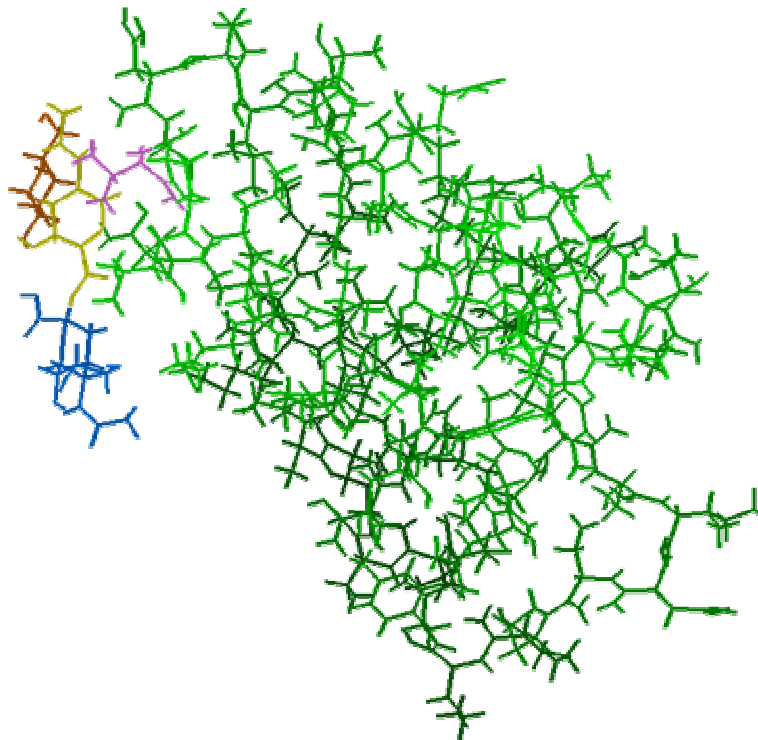


FIGURE 5: Domain III with three sugar molecules

- The structure obtained by making Domain III as small as possible by just taking twenty-one amino acids i.e. ten on either side of Threonine looks like the Figure shown below(Figure 6).

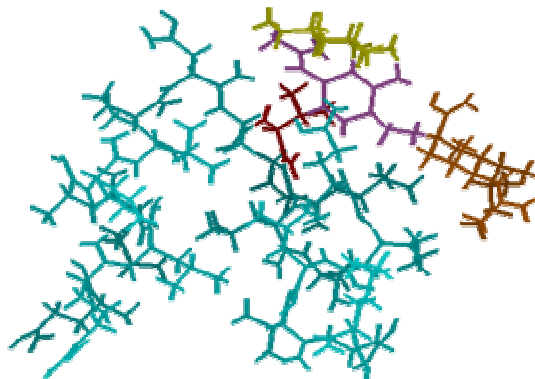


FIGURE 6: Small peptide domain with twenty-one aminoacids and three sugars

- After removing two sugars Sialic acid and Galactose the structure obtained was as seen in Figure below(Figure 7) .

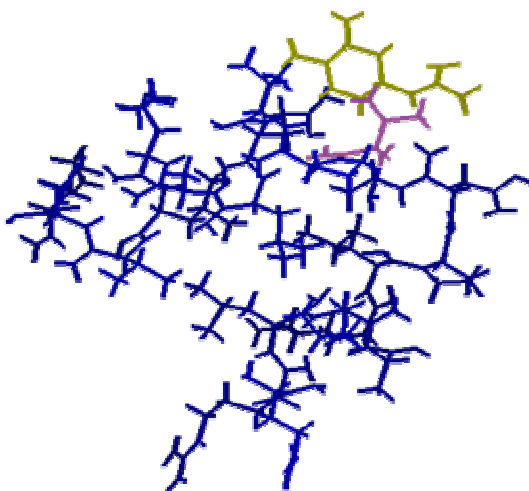


FIGURE 7: Small peptide domain with twenty-one amino acids and GalNAc

- The created linear polypeptide looks as Figure below (Figure 8). It was created using Alfa Helix. After some minimizations the structure looks like below. It consists of twenty-one amino acids chosen from the original sequence from Domain III.

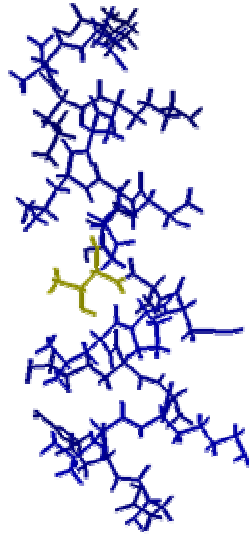


FIGURE 8: *Linear Polypeptide*

- Linear polypeptide after binding with oligosaccharide looks like the Figure shown below (Figure 9). We were able to minimize it only to some extent. Some more dynamics have to be run in order to obtain a similar structure as the small peptide domain.

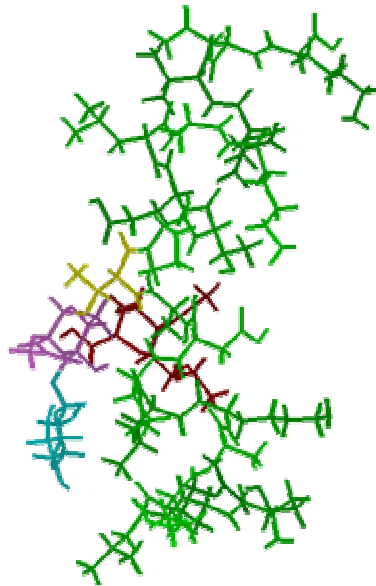


FIGURE 9: *Linear Polypeptide with three sugars*

- Linear polypeptide with GalNAc attached to Threonine looks as Figure shown below (Figure 10). We were able to minimize it only to some extent. Some more dynamics have to be run in order to obtain a similar structure as the small peptide domain.

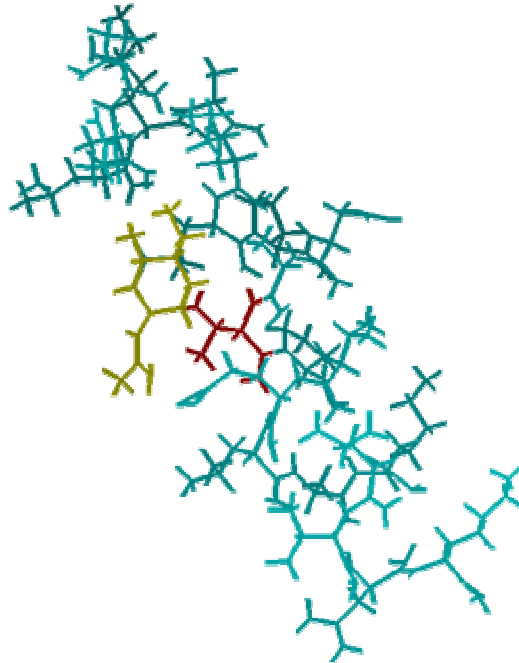


FIGURE 10: *Linear Polypeptide with GalNAc*

CONCLUSIONS:

The basic objective of our project was to model the Vitamin D protein molecule and analyze the structure under various conditions. We also wanted to manually create a linear polypeptide with 20 amino acids similar to it to the DBP Domain III molecule that had just 20 amino acids. Our intension was to observe if there was any similarity between these two structures. We successfully performed various operations using Insight II software and we obtained interesting results. We found that the two structures almost resembled each other at the present minimization values. However we need to perform more minimizations and dynamics in order to achieve the desired result. Since at higher minimizations we expect the structure to change more and embed the sugar molecule attached to it. Next we observed that the Vitamin D protein molecule that was obtained from NCBI was already in a stabilized state and we needed to perform very few minimizations to bring it to the desired global minimum. This suggests that the molecule exists in stable state in nature.

REFERENCES

1. Mariana Ramierz-Avarado, Francisco J. Blanco and Luis Serrano. 1996. *De novo* design and structural analysis of model hairpin peptide system. *Nature Structural Biology*.
2. Nobuto Yamamoto and Venkateswara R. Naraparaju. 1996. Role of Vitamin D3- Binding Protein in Activation of Mouse Macrophages. *The Journal of Immunology*. 157:1744-1749.

Project Report CBS 530

**Molecular Modeling -Comparison of
de novo predictions with known structure
(GcMAF)**

by

Vuppaladadium, NagaDeepa
Gupta, Shubhra
Ganta, Srilakshmi
Edupuganti, Vidyadhari