Chapter 6

Features of homotetrameric molecular association in the crystals of lectins and other proteins
6.1. Summary

The crystal structures of proteins with homotetrameric association, a common feature of many lectins, were analyzed to understand the characteristics of tetrameric association in terms of the arrangement of their subunits and its biological significance. The analysis could group the tetramer units into four categories:

1. Tetrahedral molecules, in which the four monomers form a nearly perfect tetrahedral arrangement. The angle between axes of any two monomers is $\sim 109^0$. Sometimes, the tetrahedral shape is distorted that is, the tetrahedral angle deviates from ideal value, which gives the molecule a twisted shape rather than a perfect tetrahedral shape.

2. Molecules that form sandwiched dimer of dimers where the two dimers are arranged perpendicular to each other, one upon the other.

3. Planar molecules where the four monomers lie in one plane and the corresponding sides of adjacent monomers face opposite directions (that makes diagonally opposite monomers to face the same direction). This can be considered a flattened tetrahedral shape.

4. Planar closed molecules where all the four monomers lie in one plane arranged in a head-to-tail fashion in a square.

The first group is the most commonly found arrangement. The importance of each arrangement for its biological function will be discussed. In addition, some unusual tetrameric assemblies were also observed, which were found to be relevant with the biological function of the particular protein.
6.2. Introduction

Many plant lectins, such as the *Artocarpus hirsuta* lectin reported from our group, concanavalin A analyzed for crystal polymorphism in the previous chapter of this thesis and other proteins such as penicillin V acylase and conjugated bile acid hydrolase also reported from our lab all showed a tetrameric association of their subunits. From the analysis of the structure of *Artocarpus hirsuta* lectin it was concluded that a pattern of subunit-association was formed to keep the functional sites of subunits at maximum distance away from each other. Obviously for a tetrameric association of single binding-site subunits a tetrahedral positioning is the most ideal one to minimize steric hindrance between ligands. This also helps to form networks of cells during agglutination of red blood cells. Based on this idea an analysis of the structures of proteins that are known to form tetramers was undertaken.

Researchers have approached the problem of protein assembly under different contexts. A variety of purposes were attributed to formation of oligomers and complexes. For example, in the case of virus assembly the genetic economy was identified as the advantage for subunit interactions and arrangement (Dokland, 2000; Phelps et al., 2000). Transcription and association of several copies from the same gene can minimize errors or alternatively, can uniformly distribute mutations throughout the assembly. For protein like hemoglobin the requirement of tetramer was for the allosteric control of oxygen binding (Monod et al., 1965). In certain membrane proteins subunit assembly helps to create an external hydrophobic and internal polar surface to help ion transport (Manting et al., 2000; Sakaguchi et al., 1997). Oligomerisation is proposed in situations where regulation of concentration levels of constituent subunits is required (Bray and Lay, 1997), which also results in increased stability and reduced surface area of the constituent molecules (Larsen et al., 1997). In certain enzymes subunits assemble to form a symmetric substrate-binding cleft such as in the HIV protease (Wlodawer and
Erickson, 1993). Oligomerisation is identified as one of the ways to achieve thermostability of proteins in thermophilic organisms (Walden et al., 2001). As may be expected, in most cases oligomerisation leads to higher order of complexity.

Many researchers have explored various aspects of protein oligomerisation and their relevance in biology (Ali and Imperiali, 2005). One study estimated more than one-third of the cellular proteins form oligomers (Goodsell and Olson, 2000). Although, oligomers can be composed of multiple subunits of the same polypeptide (homo-oligomer) or different polypeptides (hetero-oligomer), the preference seems to be for the formation of homo-oligomers in the cell (Goodsell, 1991). Similarly, even though oligomeric proteins can be formed from any number of subunits the average oligomeric state of cellular proteins estimated was tetrameric (Goodsell, 1991). In homo-oligomeric proteins, since the constituents of the assembly are identical, its formation can introduce simple point group symmetry. Goodsell and Olson (2000) have estimated that the point group symmetries of cyclic, dihedral and cubic are most frequently observed.

The stability of an oligomer will directly depend on the strength of association of the subunits, their affinity and duration. Thus, one with strong subunit-interactions can invariably be found as an oligomer, while the formation of oligomers by those with weaker interaction may depend on the concentration and other conditions like pH and temperature of the solution or in response to some other stimuli (Nooren and Thornton, 2003a, b).

Many researchers have attempted to rationalize and quantify protein-protein recognition, the type of interactions and the nature of interface involved in protein oligomerisation in a wider sense (Chothia and Janin, 1975; Miller et al., 1987; Argos, 1988; Janin et al., 1988; Miller, 1989; Jones and Thornton, 1996). Rationalization of quaternary structure formation in the case legume lectins is carried out in terms of the parameters of buried hydrophobic surface, interaction energy and shape.
complementarity at the interface of subunits (Prabhu et al., 1999). Similarly, there were attempts to study protein oligomerisation and stability in the case of legume lectins by analyzing the unfolding of their quaternary organisation (Srinivas et al., 2001). Here the attempt is to analyze the quaternary structures of tetrameric lectins and other proteins with tetrameric association in terms of their symmetry of organization and biological relevance.

6.3. Materials and Methods

All the computational work was carried out on a Silicon Graphics workstation (Octane) with Irix 6.5 as the operating system as well an IBM PC with Fedora Core 6 as the operating system. Atomic coordinates of various homotetrameric proteins were downloaded from the Protein data bank (PDB, http://www.rcsb.org/pdb/) according to their space groups, at a cut-off of sequence identity of 70%. The information in the PDB file (REMARK 350) was used to decide whether the biological unit of the protein is a homotetramer or not. Although an attempt has been made to cover all unique tetrameric protein structures in the analysis, it is likely that some of the structures might have been excluded if they were not categorized as tetramers by PDB. Hetero-octameric or sometimes, even hetero-dodecameric molecules in which one subunit consisted of two or three different monomers, and four such subunits formed the biological molecule, were also considered for the analysis purpose.

Wherever the asymmetric unit differed from the biological unit, the coordinates of the biological unit were downloaded from the PDB website (*.pdb1 files), or, sometimes the biological unit was generated by calculating the coordinates of symmetry related subunits. The protein structures were visualized using the graphical software QUANTA (Accelrys, Inc.) and grouped into various classes. The secondary structure content of the protein was also roughly estimated, to check whether there is any dependence of the
overall assembly of the molecule on the secondary structure. Centers of mass of the
tetramer and its subunits for selected structures were calculated as described in Chapter
5 (section 5.3). The diagrams of quaternary structures of proteins and that of centers of
mass were prepared using PyMOL (DeLano, 2000).

6.4. Results and Discussion

At a sequence identity cut-off of 70%, about 700 unique homotetrameric protein
structures were analyzed. The structures were grouped according to the space groups to
find any trend in preferred quaternary structure in a particular space group. These
protein structures could be grouped into four major categories.

6.4.1. Dihedral / Tetrahedral type assemblies

In this type, the four subunits of a protein are arranged pointing towards the four
corners of an approximate tetrahedron. Concanavalin A (ConA), described in chapter 5,
exhibits a near-perfect tetrahedral shape (Fig. 6.1. (A)), as shown by the measurements
of the tetrahedral angles between the centers of mass of the four subunits. Other
tetrameric legume lectins, for example *Dioclea grandiflora* lectin (PDB code 1DGL;
Rozwarski et al., 1998), *Phaseolus vulgaris* lectin (PDB code 1FAT; Hamelryck et al.,
1996) as well as hetero-octameric *Dolichos lablab* lectin (FRIL, PDB code 1QMO;
Hamelryck et al., 2000) also show a near-perfect tetrahedral arrangement of subunits. A
variation of this kind of arrangement is observed in many other proteins, including
enzymes and the arrangement gives an internal dihedral symmetry to the tetramer (point
group 222).

Most of the tetrameric proteins show two steps of oligomerisation. First, two
monomers associate to form a dimer and two dimers in turn associate to form a tetramer
(Powers and Powers, 2003). Due to this, even in the near-perfect tetrahedral shaped
molecules, the angles between any two monomers deviate at least slightly from the normal tetrahedral angle of $109.28^\circ$. The angles between two monomers of the same dimer have the least value, $\sim 100^\circ$, followed by the angles between adjacent monomers of two different dimers, which are at least $3-4^\circ$ more than the angle between monomers of the same dimer. The angles between diagonally oppositely placed subunits have values of $\sim 125^\circ$, or sometimes even more, to compensate for the other four angles.

Not all homotetrameric proteins show such a near-perfect tetrahedral shape. Most of the proteins have the angles between their subunits deviating much more, distorting the tetrahedron, and this gives the molecule a twisted shape rather than a perfect tetrahedron.

**Fig. 6.1.** (A) A perfect tetrahedral arrangement of subunits as observed in the tetramer of ConA (PDB code 1QDC). The carbohydrate ligand, man($\alpha$1-6)man($\alpha$1-O)methyl binds at the four corners of the tetrahedron. (B) Distances and angles between centers of mass of four subunits in the case of 1QDC. The distances are shown in green. Pink: angles between the subunits of the same dimer; blue: angles between the adjacent subunits of different dimers and orange: angles between the diagonally opposite subunits. Same color coding has been used for all subsequent diagrams of centers of mass.
6.4.2. Sandwiched dimer of dimers; two perpendicularly placed dimers

In this type also, two protein monomers associate to form a dimer and two dimers associate to form a tetramer. However, these dimers are placed roughly perpendicular to each other, that is, if each dimer is considered to be enclosed in a box; the two boxes appear perpendicular to each other. This was thought to be a distortion of the tetrahedral shape. The angles between monomers of the same dimer are reduced to ~80° and those between the adjacent monomers of two dimers range between 100-120°. The angles between two oppositely placed monomers have values in the range of 140 to 160°. An example of such arrangement is the indole-pyruvate decarboxylase from Enterobacter cloacae (PDB code 1OVM; Schutz et al., 2003) shown in Fig. 6.2. (A), (B) and (C). However, in another such protein, SecB from E. coli (PDB code 1QYN; Dekker et al., 2003), the angles between monomers of the same dimer are ~100° while those between the adjacent monomers of two dimers are ~80° (Fig. 6.2. (C), (D) and (E)). A schematic representation of this kind of arrangement is shown in Fig. 6.2. (G). This kind of arrangement also produces a 222 symmetry in the molecule, and it may be having the same biological significance as that of tetrahedral arrangement.
Fig. 6.2. (A) and (B) Indole-pyruvate decarboxylase from *Enterobacter cloacae* (PDB code 1OVM) and (D) and (E) SecB from *E. coli* (PDB code 1QYN) shown in two different orientations. These two proteins are dimer of dimers and the two dimers are roughly perpendicular to each other. Subunits colored in green and cyan form on dimer and those colored in magenta and yellow form another dimer. (C) and (F) Distances and angles between centers of mass of four subunits of 1OVM and 1QNY respectively. (G) Schematic representation of the dimers placed perpendicular to each other.
6.4.3. Planar assembly

This kind of arrangement is characterized by the presence of all four subunits of the protein in a single plane, with any two adjacent subunits facing in the opposite directions. This makes the diagonally oppositely placed subunits to face in the same direction. This can be considered as a flattened tetrahedral shape. From solvent accessibility calculations, it is assured that even in this type, the protein first dimerizes and the dimers associate to form tetramers. The angles between two monomers of the same dimer range from 70 to $80^0$, while those between the adjacent monomers of different dimers have values of $\sim100^0$. However, the angles between the diagonally opposite monomers have values very close to $180^0$, which are responsible for the flattened or planar shape of the molecule. This type of arrangement also shows the point group 222 symmetry between the subunits. One such example is the enzyme Penicillin V acylase from *Bacillus sphaericus* (PDB code 2PVA, Suresh et al., 1999), shown in Fig. 6.3. (A) and (B); (C) and (D) show similar kind of arrangement in TenA homolog from *Pyrobaculum aerophilum* (PDB code 2GM8), an all-alpha protein, whereas in (E), the distances and angles between centers of mass of four subunits have been shown and (F) displays a schematic representation of the molecules, showing the substrate binding sites face up.
Fig. 6.3. (A) Top view and (B) side view of the enzyme Penicillin V acylase from *Bacillus sphaericus* (PDB code 2PVA). The bound dithiane diol molecules are shown as space filling models. The ligand-binding sites of any two adjacent subunits, which are same as substrate-binding sites, lie on the opposite faces. (C) Top view and (D) side view of TenA homolog from *Pyrobaculum aerophilum* (PDB code 2GM8), an all-alpha protein. (E) Distances and angles between centers of mass of four subunits. (F) Schematic representation of the molecules, showing the substrate binding sites face up.
6.4.4. Planar closed molecules

The three types of subunit arrangements in tetrameric proteins described so far show point group 222 symmetry. The fourth kind of arrangement, in which all the four subunits lie in a plane and face in the same direction, shows a 4-fold symmetry between the subunits. Due to this, a “closed” tetramer is formed. The angles between any two adjacent monomers are nearly 90° and those between subunits oppositely placed are almost equal to the 180° and in two opposite directions. Many membrane-bound proteins, such as in ion channels and cell surface enzymes show this kind of arrangement, for example the potassium channel from *Streptomyces lividans* (PDB code 1BL8, Doyle *et al.*, 1998) which is shown in Fig. 6.4. (A) and (B); in (C) the distances and angles between centers of mass are shown; a schematic representation of this kind of arrangement is shown in (D). (E) shows a potassium channel embedded in the membrane lipid bilayer (reproduced from biop.ox.ac.uk/www/lj2001/sansom/sansom_1.jpg).
Fig. 6.4. (A) Side view and (B) Top view of potassium channel from *Streptomyces lividans* (PDB code 1BL8). The potassium ions and a water molecule have been shown as spheres. (C) Distances and angles between centers of mass of four subunits. (D) Schematic representation of the molecule exhibiting a four-fold symmetry. (E) Potassium channel in the membrane lipid bilayer. (reproduced from biop.ox.ac.uk/www/lj2001/sansom/sansom_1.jpg)
6.4.5. Tetrameric arrangements not belonging to the patterns described

Although most of the homotetrameric molecules could be grouped into one of the above mentioned four categories, some molecules could not be categorized. Most of such molecules, when analyzed in detail, were found to be wrongly labeled as homotetramers in their respective PDB files. Such molecules were discarded from further analysis. However, some of the molecules were found to be genuinely homotetrameric, still displayed a quaternary structure which could not be fitted into any of the above mentioned categories. They did not show any other recognizable pattern as well. The arrangements of subunits in some such molecules have been described below.

**Peanut lectin (PDB code 2PEL)**

This legume lectin, despite sharing sequence as well as secondary and tertiary structure similarity with other legume lectins, shows a very peculiar, “open” quaternary structure. It contains two identical dimers, each having a two-fold symmetry in its subunits, however, at the quaternary structure level; the molecule does not show any 222 or 4-fold symmetry (Fig. 6.5. (A) and (B)) (Banerjee et al., 1994). This unusual structure was also responsible for the difficulty in solving its structure (Vijayan, 2007).

![Fig. 6.5. (A) Quaternary structure of peanut lectin (pdb CODE: 2PEL). (B) Distances and angles between centers of mass of four subunits.](image-url)
**DNA binding proteins**

Several DNA binding proteins such as lactose operon repressor protein (LacR) from *E. coli* (PDB code 1TLF; Friedman *et al.*, 1995) form a very peculiar quaternary structure, which consists of two dyad-symmetric dimers which are nearly parallel to each other. Due to this, all four DNA binding domains of intact LacR are placed on the same side of the tetramer. This results in a deep, V-shaped cleft between the two dimers. An antiparallel four helix bundle which is formed from four C-terminal helices, one contributed by each monomer functions as a tetramerization domain (Fig. 6.6. (A) and (B)). While binding to DNA, the tethered dimers of this protein broaden by ~12° and the dimers twist by ~8° (Lewis *et al.*, 1996). Removing the C-terminal helix which assists in oligomerization of the protein, the affinity of the dimer remains the same (Brenowitz *et al.*, 1991); however, the induction ratios decrease (Oehler *et al.*, 1990).

![Fig. 6.6. (A)](image1) Quaternary structure of the lactose operon repressor protein (PDB code 1TLF). The C-terminal helices of all four monomers are involved in the tetramerisation of the molecule. *(B)* Distances and angles between centers of mass of four subunits.
λ phage transcription activator protein CII (PDB code 1XWR)

This is another DNA-binding protein which binds to a unique direct-repeat sequence T-T-G-C-N6-T-T-G-C, observed in three phage promoters it activates. The tetramer is formed from dimers but does not exhibit any closed symmetry (Fig. 6.7. (A)). Arrangement of centers of mass of four subunits is also peculiar for this protein (Fig. 6.7. (B)). Here also, the tetramerization is achieved by formation of a four-helix bundle, each helix contributed by a monomer. The unusual quaternary structure of this protein allows it to place the helix-turn-helix motifs of two of the four CII subunits for interaction with successive major grooves of B-DNA, from one face of DNA and helps to identify a direct repeat DNA sequence rather than the inverted repeat sequence (Datta et al., 2005).

Fig. 6.7. (A) λ phage CII protein (PDB code 1XWR). This protein also has four helices involved in tetramerisation. (B) Distances and angles between centers of mass of four subunits, depicting their quite unusual arrangement.
*Mycobacterium tuberculosis* D-3-Phosphoglycerate Dehydrogenase (PGDH)

The crystal structure of this enzyme (PDB code 1YGY, Dey et al., 2005) consists of a dimeric asymmetric unit, made of two identical subunits, each consisting of four domains. However, in one of the two subunits there is a rotation of ~180° around a hinge region connecting two of the four domains. This introduces significant asymmetry in the dimer. Two such asymmetric units associate to form a biologically active tetramer (Fig. 6.6. (A)). Distances and angles between centers of mass of four subunits are shown in (B). This asymmetric arrangement leads to the formation of two different and distinct domain interfaces between identical domains in the asymmetric unit as well as introduces asymmetry in the substrate binding sites, which might have a role in the activity and regulation of the enzyme (Dey et al., 2005).

**Fig. 6.6. (A)** Unusual quaternary structure of *M. tuberculosis* PGDH. All four subunits have identical primary structure and consist of four domains each; however, in the subunits colored in cyan and yellow, two of the four domains are flipped by almost 180° around a hinge region. This introduces the asymmetry in the dimer.

**Fig. 6.6. (B)** Distances and angles between centers of mass of four subunits, which form corners of an approximate rhombus.
6.4.6. Biological significance

In the case of most of the proteins exhibiting a tetrahedral shape, the binding sites are located at the four corners of the tetrahedron (Fig. 6.1. (A)). This reduces the steric hindrance between the ligand molecules and hence increases efficiency of the molecule. This could be the possible reason for the tetrahedral or the distorted tetrahedral shape being the most commonly observed feature of most of the homotetrameric molecules.

As seen in the case of molecules with tetrahedral shape, planar molecules which have their adjacent subunits facing in the opposite directions also have their binding sites placed at the maximum distance from each other, which reduces the steric hindrance between the ligand molecules. Enzymes belonging to this category have an additional benefit. Since their binding sites are placed on two opposite sides, substrate approaching from any side encounters the active site and hence the efficiency of the molecule increases.

The arrangement of monomers involving a four fold symmetry and hence closed planar pattern seems to be favoured by membrane bound proteins like aquaporins, plastocyanines and potassium channel proteins, or DNA binding proteins like RUVA. The reason could be that this kind of arrangement gives polarity to the molecule due to which all the hydrophobic part of the molecule gets buried in the membrane and the hydrophilic part remains exposed.
6.4.7. Correlation of subunit arrangement with the crystal systems

As the homotetrameric protein structures were also grouped according to their space groups, the prevalence of each type in particular space group was evident. As expected, the largest number of molecules displayed a tetrahedral or distorted tetrahedral shape in almost all of the crystal systems. Other types of quaternary arrangements which display the 222 symmetry, namely planar molecules with adjacent subunits facing in opposite directions were also observed almost in all crystal systems. Molecules that are sandwiched dimers of dimers, with both dimers roughly perpendicular to each other, are mainly observed in orthorhombic space groups..

In triclinic and monoclinic space groups, the asymmetric unit is always a tetramer, or sometimes, even two tetramers. In orthorhombic space groups, monomeric or dimeric asymmetric units are also observed which are located at special position and the crystallographic symmetry operations give rise to the functional tetramer. Occurrence of monomeric/dimeric asymmetric units is also common in space groups with two 2-fold axes belonging to tetragonal and hexagonal crystal systems, such as \( P4_{x22} \) and \( P6_{x22} \), where “x” denoted the screw axis. Trigonal space groups with a 2-fold axis, such as \( P3_{x21} \) or \( P3_{x12} \) may show presence of dimeric asymmetric unit placed at the special position. All these conditions give rise to a 222 symmetry in the molecule.

Planar closed arrangement of subunits involving a 4-fold symmetry is relatively rare. This type is mainly observed in the tetragonal or cubic space groups, where in most cases, the asymmetric unit is a monomer and the biological tetrameric molecule can be generated using the symmetry operations. Rarely, this kind of arrangement is seen in orthorhombic or monoclinic space groups as well, but then the asymmetric unit is always a tetramer.

Table 6.1 enlists all crystal systems with the prevalent types of quaternary arrangements observed in each.
Table 6.1. The seven crystal systems with different tetrameric arrangements observed in them.

<table>
<thead>
<tr>
<th>Crystal System</th>
<th>Symmetry in tetramer</th>
<th>No. of molecules in asymmetric unit</th>
<th>Prevalent Type of subunit arrangement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triclinic</td>
<td>Mostly 222</td>
<td>4</td>
<td>Tetrahedral/Distorted tetrahedral</td>
</tr>
<tr>
<td>Monoclinic</td>
<td>Mostly 222</td>
<td>4, 8</td>
<td>Tetrahedral/Distorted tetrahedral</td>
</tr>
<tr>
<td>Orthorhombic</td>
<td>Mostly 222</td>
<td>1, 2, 4, 8</td>
<td>Tetrahedral/Distorted tetrahedral</td>
</tr>
<tr>
<td>Trigonal</td>
<td>222</td>
<td>4 in space groups without 2-fold axes, 2, 4 in space groups with 2-fold axes</td>
<td>Tetrahedral/Distorted tetrahedral</td>
</tr>
<tr>
<td>Hexagonal</td>
<td>222</td>
<td>4 in space groups without 2-fold axes, 1, 2, 4 in space groups with 2-fold axes</td>
<td>Tetrahedral/Distorted tetrahedral</td>
</tr>
<tr>
<td>Tetragonal</td>
<td>222 or 4-fold</td>
<td>1, 2, 4</td>
<td>Planar 2 #</td>
</tr>
<tr>
<td>Cubic</td>
<td>222 or 4-fold</td>
<td>1, 2, 4</td>
<td>Planar 2</td>
</tr>
</tbody>
</table>

* Planar 1: Planar molecules in which adjacent subunits face in the opposite directions; exhibiting 222 symmetry.

* Planar 2: Closed planar molecules displaying 4-fold symmetry.
6.5. Conclusions

Study of unique homotetrameric protein structures reported in PDB revealed four types of prevalent arrangements of subunits in the tetramer. Three of these four types showed point group 222 symmetry in their subunits, while the fourth type exhibited a 4-fold symmetry. While the 222 symmetry is commonly observed in many proteins including lectins and most of the enzymes, the four-fold symmetry is restricted to most of the membrane bound proteins like ion channels and certain membrane bound enzymes. Some other unusual quaternary structures were also observed in this study, which did not form any group of their own. Such unusual arrangements could be correlated with the specific biological activity of the concerned protein.