CRYSTAL STRUCTURE DETERMINATION OF HEX1 AND HUMAN GEMININ70-152

YUAN PING

NATIONAL UNIVERSITY OF SINGAPORE

2004

Founded 1905

CRYSTAL STRUCTURE DETERMINATION OF HEX1 AND HUMAN GEMININ70-152

YUAN PING

(*B.S., M.Sc.*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

INSTITUTE OF MOLECULAR AND CELL BIOLOGY

NATIONAL UNIVERSITY OF SINGAPORE

2004

ACKNOWLEDGEMENTS

I want to express my sincere and deep gratitude to my supervisor Dr. Kunchithapadam Swaminathan for his expert guidance, encouragement and support to undertake and finish my Ph.D. study successfully.

Next, I wish to express my special thanks to Prof. Anindya Dutta (Byrd Professor of Biochemistry & Molecular Genetics, Professor of Pathology, University of Virginia Health Sciences Center) for initiating the project of structure determination of Geminin and generously allowing me to include the functional data in my thesis to get the story complete. Besides, he also supported me to work for half a year in his previous lab at Brigham and Women's Hospital, Harvard Medical School. There I started to learn DNA replication and exposed myself to a world of first class research that greatly enlarged my vision. I also wish to thank Drs. James Wohlschlegel, Zophonias O. Jonsson, Yuichi Machida, Suman Kumar Dhar, Sandeep Saxena, and other members in his lab for their kind assistance and valuable training in molecular biology.

I owe my sincere thanks to our collaborators Prof. Nam-Hai Chua and Dr. Gregory Jedd (Laboratory of Plant Molecular Biology, The Rockefeller University) for initiating the project of crystal structure determination of Hex1 and exchanging data during the collaboration.

I wish to express my sincere gratitude to Prof. Subramanyam Swaminathan, Drs. Desigan Kumaran, and Howard Robison (Department of Biology, Brookhaven

National Laboratory) for their great help in data collection and valuable suggestions in structure determination.

I am very glad to express my sentiments to my colleagues of the structure lab, especially Ms. Nan Li, Sifang Wang, and Sheemei Lok, who created a friendly atmosphere.

Here I also want to express my deep thanks to my parents who raised me up and encouraged me to pursue the doctoral degree.

Finally, I want to express my greatest love and regards to my husband Mr. Yifeng Sun for his full support and encouragement. Without his help, I will not be able to finish my Ph.D. and have such a happy life. Here, I dedicate this thesis to the crystal of our love - our son, Ruiqian Sun.

TABLE OF CONTENTS

NOMENCLATURE

X-ray crystallography, unit-cell, lattice, point group, space group, hkl plane, reciprocal lattice, electron density map, Fourier series, Fourier transformation, data reduction, anomalous scattering, MIR, MAD, MR, direct method, refinement

Hex1, Woronin body, eIF-5A

DNA replication, pre-replication complex, neuralization, apoptosis, leucine zipper, coiled coil

LIST OF FIGURES

LIST OF TABLES

SUMMARY

X-ray crystal structure determination is one of the most powerful methods to determine the macromolecular structure and study the relationship between structure and function of macromolecules.

With this method, I have solved the crystal structure of Hex1, the component of Woronin body in *Neurospora crassa,* at 1.8 Å by the MAD method. The Woronin body is a dense-core vesicle specific to filamentous ascomycetes where it functions to seal the septal pore in response to cellular damage. Previous work showed that the Hex1 protein self-assembles to form the solid core of the Woronin body. The structure of Hex1 reveals the existence of three intermolecular interfaces that promote the formation of a three-dimensional protein lattice. Point mutation of the intermolecular contact residues and expression of an assembly-defective Hex1 mutant result in the production of aberrant Woronin bodies, which possess a soluble non-crystalline core. This mutant also fails to complement Hex1 deletion in *Neurospora crassa*, demonstrating that the Hex1 protein lattice is required for Woronin body function. In addition to sharing sequence similarity, the tertiary structure of Hex1 is remarkably similar to that of eukaryotic initiation factor 5A (eIF-5A). Thus it suggests that a new function of Hex1 has evolved following the duplication of an ancestral *eIF-5A* gene, which may define an important step of fungal evolution.

With the X-ray crystallography method, I have also solved the crystal structure of the Cdt1 binding domain of Geminin at 2.0 Å. For a cell to survive, the chromosome of the cell should be accurately replicated only once and then equally

divided to two daughter cells. Very subtle biological switches control these two steps. Geminin plays an essential role in controlling the chromosome to replicate only once. Before DNA replication, a pre-replication complex (pre-RC) should be formed in a stepwise manner. Origin recognition complex (Orc) is always associated with the chromatin DNA origin. When cells exit mitosis and enter G1, Cdc6 and Cdt1 will load on Orc. Then microchromosome maintenance (MCM) complex will join in to form pre-RC and it is the mark of the beginning of S phase. Only the DNA that has been loaded with pre-RC, called "licensed", can be replicated. To achieve the target that DNA replicates only once during S phase, Geminin will function to ensure that no new pre-RC is formed on an already fired origin. By interacting with Cdt1, Geminin targets Cdt1 for degradation. Without Cdt1, MCM will not be able to load to chromatin. Thus no PreRC will be formed and a fired origin will not be fired again. Residues 70-152 is the functional domain of Geminin that can interact with Cdt1 and also inhibit EBV oriP based transient plasmid replication. The crystal structure of Geminin70-152 clearly reveals amino acids 92 to 152. Amino acids from 70 to 91 are missing in the electron density map, which suggests the region may be highly flexible. The fragment from amino acid 94 to 150 forms dimerized parallel coiled coil structure. This indicates that full length Geminin also forms a dimer. Point mutations of leucine and isoleucine residues in the coiled coil domain disrupt dimerized Geminin and also abolish its interaction with Cdt1 *in vitro* and *in vivo*. This mutant also loses its ability to inhibit EBV plasmid replication as well as DNA replication in *Xenopus* egg extract. These data show the coiled coil structure of Geminin is critical for its interaction with Cdt1 and inhibits DNA replication. Further experimental data reveal that Geminin 93-152 is sufficient for Cdt1 interaction, but not sufficient for inhibiting Cdt1's function. Residues 70 to 92 appear to be necessary to inhibit DNA replication. The physical

contiguity of residues 70 to 93 with the coiled coil domain might indicate that the critical function of this accessory domain may either stabilize the interaction of Cdt1 with Geminin or make additional contacts with Cdt1 that interfere with whatever function is necessary for co-operating with Cdc6 to load the Mcm2-7 helicases. Alternatively, this domain of Geminin might have a novel interaction partner. Further work on the structure of the Geminin-Cdt1 complex may give us more hints of the mechanism.

CHAPTER 1 INTRODUCTION ON CRYSTAL STRUCTURE DETERMINATION

1.1 THE HISTORY OF X-RAY CRYSTALLOGRAPHY

1.1.1 Discovery of X-rays

Discovered by German physicist Wilhelm Conrad Roentgen in 1895, X-rays lie in the electromagnetic spectrum between ultraviolet and gamma radiation and have wavelengths of 0.1-100 Å. They are usually produced by rapidly decelerating fast moving electrons and converting their energy of motion into quanta of radiation (Stout, 1989). When high energy electrons collide with and displace an electron from a low lying orbital in a target metal atom, an electron from a higher orbital drops into the resulting vacancy, emitting its excess energy as an X-ray photon (Rhodes, 2000). The wavelengths (λ) of emission lines are longer for elements of lower atomic number Z. For instance, electrons dropping from the L shell of copper (Z = 29) to replace the displaced K electrons (L to K or K_α transition) emit X-rays of λ = 1.54 Å. The M$\rightarrow$K transition produces a nearby emission band (K_β) at 1.39 Å. For molybdenum (Z = 42), λ (K_α) = 0.71 Å and λ (K_β) = 0.63 Å.

1.1.2 Application of X-rays to molecular structure determination

In 1912, Von Laue's group discovered X-ray diffraction and this discovery

gave rise to the development of a very rich scientific period and created a new academic branch – crystal structure determination. One year later, W. L. Bragg determined the first structure. From then on, crystal structure determination is broadly undertaken on inorganic and organic molecules (Buerger, 1990). Currently, there are about 17,000 unique structures of protein, peptide, virus, protein/nucleic acid complex, nucleic acid and carbohydrate molecules, all determined by X-ray crystallography, and deposited in the Protein Data Bank. This number is increasing day by day. Crystal structure determination is definitely the most popular and powerful method to solve a macromolecular structure.

1.2 X-RAY SOURCES AND DIFFRACTION INSTRUMENTS

1.2.1 X-ray sources

The most commonly used X-rays are produced by bombarding metals (copper or molybdenum) with electrons produced by a heated filament and accelerated by an electric field.

A monochromatic (single wavelength) source of X-rays is desirable for crystallography. Mostly, copper and molybdenum are used as the anode material and they generate suitable X-rays. Generally, the weaker K_β radiation is removed. Copper radiation is widely used for detection by film, which is more sensitive to Cu K_α than to that of molybdenum. Molybdenum, which gives short wavelength X-rays, and hence better resolution, are more commonly used when X-rays are detected by scintillation counters, as in diffractometers.

There are three common X-ray sources: X-ray tube, rotating anode tube, and the particle accelerator that produces synchrotron radiation in the X-ray region. In the X-ray tube, electrons from a hot filament (near the cathode) are accelerated by an electrical field and collide with a water cooled anode, which is made of the target metal. Output from the X-ray tube is limited by the amount of heat, which can be dissipated from the anode by circulating water. Rotating anode tubes will generate high X-ray output, in which the target is a rapidly rotating metal block. This arrangement improves heat dissipation by spreading the electron bombardment over a much larger area of metal.

Particle accelerators, which are used by physicists to study subatomic particles, are the most powerful X-ray sources. In these giant rings, electrons or positrons circulate at velocities near the speed of light, driven by energy from radio frequency transmitters, and maintained in circular motion by powerful magnets. A charged body like an electron emits energy (synchrotron radiation) when forced into circular motion, and in accelerators, the energy is emitted as X-rays. Accessory devices called "wigglers" cause additional bending of the beam, thus increasing the intensity of radiation. Systems of focusing mirrors and monochromators that are tangential to the storage ring provide powerful monochromatic X-rays at selectable wavelengths. With this principle, the X-ray data that require several hours of exposure using a rotating anode source can often be obtained in a few seconds or minutes at the synchrotron source. Another advantage is that the wavelength of synchrotron X-rays can be selected as needed, which can help to solve the phase problem.

1.2.2 Diffraction instruments

Any X-ray diffraction instrument consists of two main parts: a mechanical part for rotating the crystal and a detecting device to measure the position and intensity of diffracted beams.

Formerly, Buerger's precession camera and Weissenberg's rotation camera were used for collecting data. In both types of camera, X-ray film was used as the detector. The precession camera has the advantage of giving an undistorted image of the reciprocal lattice. Unit-cell dimensions and symmetry in the crystal can be easily derived from the undistorted image. However, the precession camera is not suitable for three-dimensional X-ray data collection, because it only records one reciprocal layer per exposure. The rotation camera registers data more efficiently, but recognition of diffraction spots is more complicated. Data are collected in a contiguous oscillation range. Some spots appear partly on one and partly on the next or previous exposures. Two "partials" are treated as individual reflections and their intensities are added at a later stage. Nevertheless, accuracy in the intensity of reflections is sometimes lower than that of fully recorded reflections and for this reason they are completely neglected. Superior resolution can be obtained from the fine grain of the photographic film, but because processing the film is somewhat cumbersome and time consuming, it is not very much used in crystallography nowadays.

The third class of instrument is a computer controlled diffractometer, which has a single counter, normally a scintillation counter. Although the scintillation counter is no longer widely used for macromolecular data collection today, the goniostat construct is standard in X-ray crystallography. The classical Eulerian geometry goniostat has four circles. The X-ray beam, the counter and the crystal lie in a

horizontal plane. The crystal is located at the intersection of the circles. To measure the intensity of a diffracted beam, the crystal must be oriented such that the diffracted beam will also be in the horizontal plane. This orientation is achieved by the rotation of the crystal around three axes: ϕ (phi), ω (omega), and χ (chi). The counter can be rotated in the horizonal plane around the 2θ-axis, which is coincident with, but independent of the ω-axis. Data collection is done either with the ω and 2θ axes coupled or with the 2θ axis fixed and the crystal scanned by rotation around the ω axis. In a kappa geometry goniostat, the equivalent rotation of the crystal is achieved by the three axes ϕ, ω and κ (kappa).

Lately, the classical photographic film has been replaced completely by the introduction of much faster electronic area detectors and image plates. The basic difference between the area detector and photographic film is that area detectors scan through a diffraction spot in small intervals (e.g., 0.1°), giving a three-dimensional profile of the spot. Area detectors are currently based on either a gas filled ionization chamber or an image intensifier, coupled to a video system. The gas filled chambers are essentially single photon counting X-ray detectors. In the video based area detectors, the diffraction pattern is collected on a fluorescent screen. In another kind of area detector the video tube is replaced by a charge coupled device (CCD). They have a high dynamic range, combined with excellent spatial resolution, low noise and a high count rate. The image plate is used in the same manner as the X-ray film but it has several advantages. The image plate is made by depositing a thin layer of inorganic storage phosphor on a flat base. Image plates are more sensitive than the X-ray film and their dynamic range is much wider (1: 10^4 – 10^5). The plate can be erased by exposure to intense white light and used repeatedly. It is also highly sensitive to the

synchrotron radiation at shorter wavelengths (around 0.65 Å; an advantage of short wavelengths is that the absorption of the X-ray beam in a protein crystal becomes negligible and no absorption correction is required). The disadvantage of the image plate is, like the photographic film, it requires a multistep process: exposure as the first step and reading as the second step (Drenth, 1994).

1.2.3 Data reduction

The goal of data collection is to collect a set of consistently measured and indexed intensities for as many reflections as possible. The preliminary manipulation of these intensities is to convert them to a corrected and more generally usable form. This process is named as data reduction. The most important quality derived from the intensities is structure factor (structure amplitude $|F_{hkl}|$), with the relationship $|F| \propto I^{1/2}$.

Because of viability in the diffracting power of crystals, intensity of the X-ray beam, slow deterioration of the crystal during data collection, and also data maybe obtained from more than one crystal, it can not be assumed that absolute intensities are consistent from one block of data to the next. An obvious way to obtain this consistency is to compare reflections of the same index that were measured from more than one crystal and scale the intensities of the two blocks of data so that identical reflection are given identical intensities. This process is called scaling. Scaling will help generate a list of internally consistent intensities for most of the available reflections (Rhodes, 2000; Stout, 1989).

1.3 BASIC CONCEPTS OF X-RAY CRYSTALLOGRAPHY

1.3.1 Unit-cell

When molecular substances enter the crystalline state from solution, individual molecule adopts one or only a few orientations. Although imaginary, the smallest box of definite volume and actual shape can be assumed to pack the molecule or all specifically orientated molecules. The crystal is an orderly three-dimensional repetition of this box. In crystallography, this identical box is defined as unit-cell (Buerger, 1966).

The dimensions of a unit-cell are designated by six parameters, three lengths of the edges a, b and c, and three unique interaxial angles α, β and γ. A cell in which $a \neq b \neq c$ and $\alpha \neq \beta \neq \gamma$ is known as triclinic while $a = b = c$, $\alpha = \beta = \gamma = 90°$ is cubic. Other unit-cells, monoclinic, orthorhombic, tetragonal, rhombohedral and hexagonal, will have a varying degree of arrangements for these parameters. The arrangement of symmetry elements (a crystal system) demands a particular unit-cell. If the unit-cell of one crystal system has dimensions that mimic a different system, it is purely accidental.

1.3.2 Lattice, point group and space group

For many geometrical purposes, it is convenient to ignore the specific nature of the motif and concentrate attention on the geometry of repetition. In these cases it is sufficient to consider using a point at the corners or vertices of the unit-cell to represent the whole unit-cell. The array of these points is called lattice. There are 14 basic unit-cells in three dimensions, called Bravais lattices. The Bravais lattices are the distinct lattice types which, when repeated can fill the whole space. They can be classified as primitive (simple unit-cell), face centered (point at the center of each face),

body centered (point at the center of the cell), end centered (point at the center of one face) and rhombohedral. Cubic cells can have primitive, body centered and face centered lattice; tetragonal cells can have primitive and body centered lattice; orthorhombic cells can have primitive, face centered, body centered and end centered lattice; hexagonal cell have primitive and rhombohedral lattice; monoclinic cells have primitive and end centered lattice while triclinic cells can only have primitive lattice.

A molecule follows certain symmetry operation when it is packed into a crystal. Beside unit translations along the three unit-cell axes, called three-dimensional translation symmetry, other symmetry elements are rotation, reflection, and inversion. The combination of these symmetry elements that acts on a unit-cell is commonly called crystallographic point group. The simplest point groups are composed of proper rotations of the individual n-fold symmetry axis. These are the point groups 1, 2, 3, 4, and 6. The total number of crystallographic point groups involving proper rotations is 11. Point groups also contain improper rotation, which is conformed to one of the six general types: $\bar{n}$, $n/\bar{n}$, *PII, IPI, IIP,* and *P/I P/I P/I*. There are 21 improper rotations. Thus totally, there are 32 crystallographic point groups (Buerger, 1956).

Rotation or reflection combined with translation will generate screw or glide symmetry, respectively. The combination of points groups, their allowed screw axes and glide planes and lattices leads to 230 different ways to combine the allowed symmetry operations in crystal, leading to 230 space groups. They were derived almost simultaneously by Fedorov (1890) in Moscow, Schoenflies (1891) in Göttingen, and Barlow (1894) in London (Buerger, 1990), and tabulated in the International Tables of Crystallography (Hahn, 1998). Because only L-amino acids are present in proteins, the application of mirror planes and inversion centers would change an L-amino acid to a

D-amino acid, not all 230 space groups are allowed and only 65 space groups are applicable to protein crystals (McRee, 1999a).

1.3.3 hkl plane

A convenient way to study the crystalline lattice is to use hkl planes. The index h gives the number of planes in the set per unit-cell in the X direction, or equivalently, the number of parts into which the set of planes cuts the a edge of each cell. Similarly, the indices k and l specify how many such planes exist per unit-cell in the Y and Z directions. The family of planes having indices hkl are the (hkl) planes.

1.4 THE DIFFRACTION OF X-RAYS BY CRYSTALS

1.4.1 Scattering by atoms in a crystal

Scattering is interaction between X-rays as electromagnetic waves and electrons. If an electromagnetic wave is incident on a system of electrons, the electric and magnetic components of the wave exert force on the electrons. This causes the electrons to oscillate with the same frequency as the incident wave. The oscillating electrons act as radiation scatterers and they emit radiation of the same frequency as the incident radiation (Drenth, 1994).

The electron cloud of an atom scatters an X-ray beam. The scattering is dependent on the number of electrons and their positions in the cloud. The atomic scattering factor can be written as a function of $2(\sin\theta/\lambda)$.

When a series of wavefronts of X-rays impinges on a row of regularly spaced

atoms, each atom produces a new set of spherical wave envelopes around itself, and any line-up of envelopes constitutes a combined wave moving in the direction of the common tangent. The cooperative combination of scattered wavelets is known as diffraction.

The combined wave scattered by the crystal can be described as a summation of the enormous number of waves, each scattered by one electron in the crystal. This may sound intimidating, because a single unit-cell in a protein crystal contains approximately 100,000 or more electrons, and there are many unit-cells in a crystal. All these waves must be added. With mathematical deduction, the calculation of diffracted waves can be simplified.

1.4.2 Waves and addition

$E = A \cos \omega t$ is normally used to represent an origin wave. A is the amplitude of the wave and ω is related to the frequency of the wave. A new wave with the same λ and same amplitude but displaced over a distance Z with respect to the original wave is described as $E_{new} = A \cos (\omega t + \alpha)$ where α is a phase shift and Z corresponds to the phase shift $2\pi (Z/\lambda) = \alpha$. Mathematically:

$$A \cos (\omega t + \alpha) = A \cos\alpha \cos\omega t - A \sin\alpha \sin\omega t = A \cos\alpha \cos\omega t + A \sin\alpha \cos(\omega t + 90)$$

So the wave $A \cos (\omega t + \alpha)$ can be regarded as being composed of two waves: wave 1 of amplitude $A \cos\alpha$ and phase angle 0° and wave 2 of amplitude $A\sin\alpha$ and phase angle of 90°. Wave 1 is called the real part and wave 2 the imaginary part of the total wave. This can be represented conveniently in an axial system called the Argand diagram, in which the real axis is horizontal and the imaginary axis is vertical. To add

several waves with different phase angles, add all the real parts and all the imaginary parts and it is the same as adding several wave vectors together (Drenth, 1994).

1.5 BRAGG'S LAW

1.5.1 Bragg's law

In Bragg's model of diffraction as reflection from parallel sets of planes, any of these sets of planes can be the source of one diffracted X-ray beam. Bragg showed that a set of parallel planes with indices hkl and interplanar spacing d_{hkl} produces a diffracted beam when X-rays of wavelength λ impinge on the planes at an angle θ and are reflected at the same angle, only if θ meets the condition

$$2\, d_{hkl} \sin\theta = n\lambda. \qquad (1.1)$$

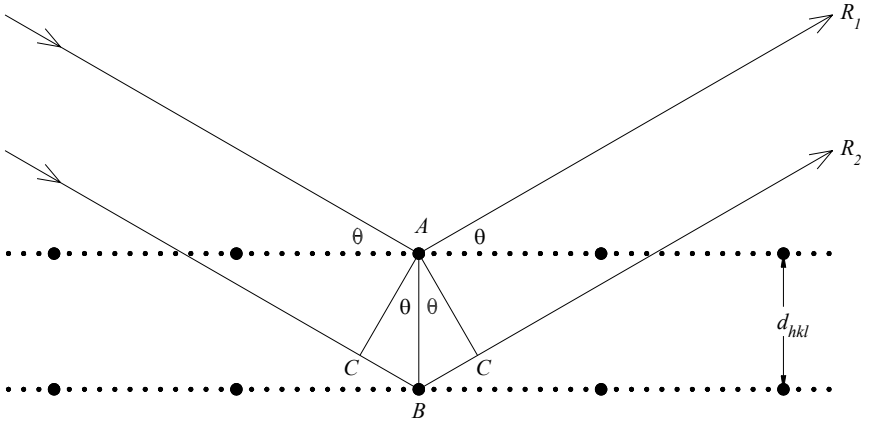


Figure 1-1. The condition that produces diffracted rays. sin θ = BC/AB, BC = AB sin θ = d_{hkl} sin θ. If the additional distance (2BC) travelled by the more deeply penetrating ray R_2 is an integral multiple of λ, then

rays R_1 and R_2 interfere constructively.

Notice that the angle of diffraction θ is inversely related to the interplanar spacing d_{hkl} ($\sin\theta$ is proportional to $1/d_{hkl}$). This implies that large unit-cells, with large spacing, give small angles of diffraction and hence produce many reflections that fall within a convenient angle from the incident beam. On the other hand, small unit-cells give a large angle of diffraction, producing fewer measurable reflections. In a sense, the number of measurable reflections depends on how many reflections are present in the unit-cell.

Each set of parallel planes in the crystal produces one reflection. The intensity of a reflection depends on the electron distribution along the planes that produce the reflection.

1.5.2 Reciprocal lattice

The diffraction pattern consists of reflections (spots) in an orderly array on the film and it has a simply inverse relationship with the spacing of unit-cells in the crystalline lattice, so the spacing of reflections in the lattice on the film is called reciprocal lattice. A reciprocal lattice is generated as explained in Figure 1-2. Take *O* as the origin, through a neighbouring crystal lattice point N, draw one plane each of the set (110), (120) and so forth, whose interplanar distances will be d_{110}, d_{120} and so on. From the origin, draw a line normal to the (110) plane. The point at a distance, $1/d_{110}$, on this line will define the reciprocal lattice point 110. Do the same for (120) and so on. Note that the points defined by this operation form a lattice, with the chosen origin. This new lattice is the reciprocal lattice. If the real unit-cell angles α, β and γ are 90°, the reciprocal unit-cell has axes *a** lying along the real unit-cell edge with the

corresponding length of 1/*a*. Similarly, the other parameters, *b** and *c** are defined. If the axial lengths are expressed in Angstroms, then the reciprocal lattice spacing is in the unit 1/Å or $Å^{-1}$ (reciprocal Angstroms).

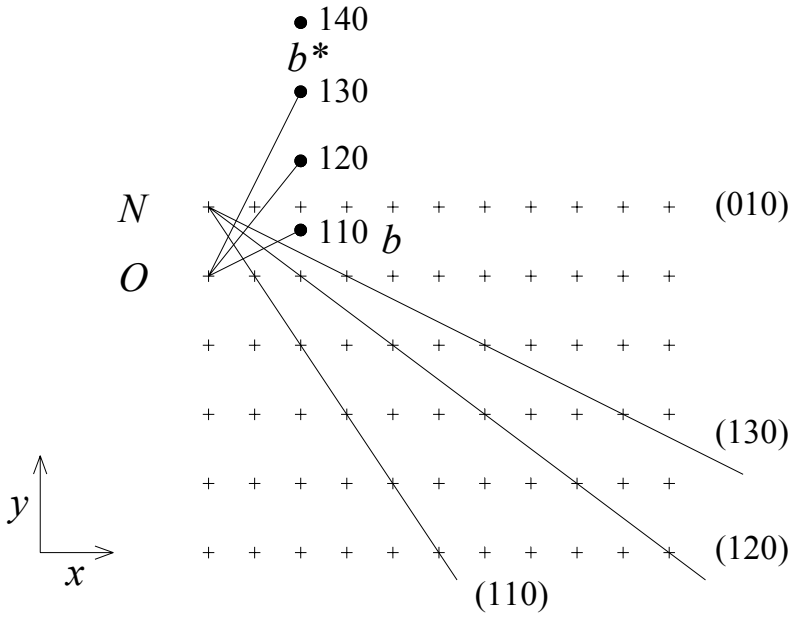


Figure 1-2. The reciprocal lattice

1.5.3 Bragg's law in reciprocal lattice

Reciprocal lattice points give the crystallographer a convenient way to compute the direction of diffracted beams from all sets of parallel planes in the crystalline lattice (real space). The following geometrical interpretation of diffraction was formulated by Ewald.

Assume that an X-ray beam (arrow *XO* in Figure 1-3) impinges on the crystal on a plane. Point *O* is arbitrarily chosen as the origin of the reciprocal lattice. *O* is also the real lattice origin in the crystal. Draw a circle of radius 1/λ with its center *C* on *XO*

and passes through *O*. This circle represents the wavelength of X-rays in the reciprocal space. Rotating the crystal about *O* will rotate the reciprocal lattice about *O*, successively bringing the reciprocal lattice points *P* and *P′* into contact with the circle. Because the triangle PBO is inscribed in a semicircle, it is a right angled triangle and $\sin\theta = OP/BO = OP/(2/\lambda)$.

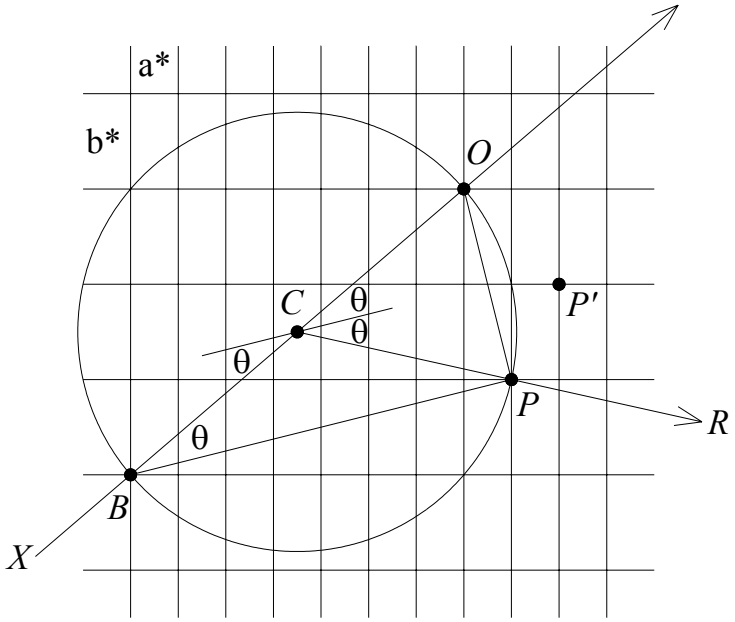


Figure 1-2. Ewald sphere

Because *P* is a reciprocal lattice point, the length of line *OP* is $1/d_{hkl}$, where *h, k* and *l* are the indices of the set of planes represented by *P*. So, $1/OP = d_{hkl}$ and $2d_{hkl}\sin\theta = \lambda$, which is Bragg's law with $n = 1$.

The line defining a reciprocal lattice point is normal to the set of planes having the same indices as the point. *BP*, which is perpendicular to *OP*, is parallel to the planes that are producing the reflection *P* in Figure 1-3. If we draw a line parallel to

BP and passing through *C*, the center of the circle, this line represents a plane in the set that reflects the X-ray beam under these conditions. The beam impinges on this plane at an angle θ, reflected at the same angle and diverges from the plane at *C* by an angle 2θ, which takes it precisely through the point *P*. *CP* gives the direction of the reflected ray *R*. In conclusion, reflection occurs in the direction *CP* when the reciprocal lattice point *P* comes in contact with this circle. As the crystal is rotated in the X-ray beam, all reciprocal lattice points come into contact with this sphere. Each reciprocal lattice point produces a beam in the direction of a line from the center of the sphere of reflection through the reciprocal lattice point that is in contact with the sphere.

This model of diffraction also implies that the directions of reflection, as well as the number of reflections, depend only on the unit-cell dimensions, and not on the contents of the unit-cell.

1.6 FOURIER TRANSFORM

1.6.1 Fourier series

Each reflection is the result of diffraction from atoms in the unit-cell. As a wave is periodic, according to the Fourier theory, any periodic function can be approximated by a series of cosine and sine terms with appropriate coefficients.

$$f(x) = \sum_{h=0}^{n} | F_h | [\cos 2\pi(hx) + i \sin 2\pi(hx)] \tag{1.1}$$

Here $f(x)$ specifies the resulting diffracting wave and it is the sum of n Fourier terms or diffraction from n atoms. Each term is a simple wave with its own amplitude $|F_h|$, its

own frequency h, and implicitly, it is own phase α_h. Since

$$cos\theta + isin\theta = e^{i\theta} \tag{1.2}$$

the above Fourier series can be written as

$$f(x) = \sum_h |F_h| e^{2\pi i(hx)} \tag{1.3}$$

When the above Fourier series is derived as a three dimensional Fourier series, the equation will be

$$f(x,y,z) = \sum_h \sum_k \sum_l |F_{hkl}| e^{2\pi i(hx+ky+lz)} \tag{1.4}$$

Here each term in the series is a simple three-dimensional wave whose frequency is h in the X direction, k in the Y direction and l in the Z direction. For each possible set of value h, k and l, the associated wave has an amplitude $|F_{hkl}|$.

1.6.2 The Fourier transform: general features

Fourier demonstrated that for any function *f(x)*, there exists another Function *F(h)* such that

$$F(h) = \int_{-\infty}^{+\infty} f(x) e^{2\pi i(hx)} dx \tag{1.5}$$

Where *F(h)* is called the Fourier transformation *(FT)* of *f(x)*, and the unit of the variable *h* is the reciprocal of the unit of *x*.

The Fourier transform operation is reversible. That is, the same mathematical operation that gives *F(h)* from *f(x)* can be carried out in the opposite direction to give

f(x) from *F(h)*, if x and h are reciprocal to each other.

$$f(x) = \int_{-\infty}^{+\infty} F(h) e^{-2\pi i(hx)} dh \tag{1.6}$$

The above functions *f(x)* and *F(h)* are one-dimensional. If stated in three dimensions, the Fourier transform would be:

$$F(h,k,l) = \int_x \int_y \int_z f(x,y,z) e^{2\pi i(hx+ky+lz)} dxdydz \tag{1.7}$$

and in turn the reverse Fourier transform is

$$f(x,y,z) = \int_h \int_k \int_l F(h,k,l) e^{-2\pi i(hx+ky+lz)} dhdkdl \tag{1.8}$$

1.6.3 Electron density as a Fourier series

The unit-cell can be represented as an assembly of electron density in several defined volume elements. The electron density of each volume element centered at (*x*, *y*, *z*) is roughly the average value of $\rho(x, y, z)$ in that region. Smaller the volume elements, the more precisely these averages approach the correct value of $\rho(x, y, z)$ at all points. In this way, the structure factor F_{hkl} can be written as:

$$F_{hkl} = \int_h \int_k \int_l \rho(x,y,z) e^{2\pi i(hx+ky+lz)} dxdydz \tag{1.9}$$

The electron density is, in turn, the transform of structure factors:

$$\rho(x,y,z) = \frac{1}{v} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)} \tag{1.10}$$

1.6.4 Structure factor as a Fourier series

The structure factor is the resultant of n waves scattered in the direction of the reflection *hkl* by the N atoms in the unit-cell. Each of these waves has an amplitude proportional to f_j, the scattering factor of the atom, and a phase α_j with respect to the origin of the unit-cell.

Crystallographers represent each structure factor as a complex vector. The length of this vector represents the amplitude of the structure factor F_{hkl}, which is proportional to $(I_{hkl})^{1/2}$. The phase is represented by the angle α that the vector makes with the real axis when the origin of the vector is placed at the origin of the complex plane. The structure factor F can be represented as a vector $A + iB$ on this plane, Figure 1-4. The projection of F on the real axis is its real part A, a vector of length $|A|$ and the projection of F on the imaginary axis is its imaginary part iB, a vector of length $|B|$.

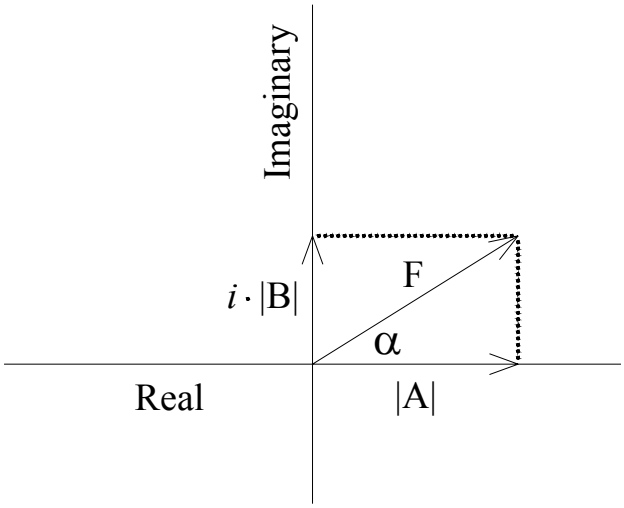


Figure 1-3. Real and imaginary components of the structure factor

From above figure:

$$\sin\alpha = \frac{|B|}{|F|} \quad \text{and} \quad \cos\alpha = \frac{|A|}{|F|} \tag{1.11}$$

Therefore,

$$|A| = |F|\cos\alpha \quad \text{and} \quad |B| = |F|\sin\alpha \tag{1.12}$$

$$F = |A| + i|B| = |F|(\cos\alpha + i\sin\alpha) \tag{1.13}$$

Expressing the complex terms in parentheses as an exponential,

$$F = |F| \cdot e^{i\alpha} \tag{1.14}$$

Substituting this expression for F_{hkl} in equation 1.14 will generate

$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l |F_{hkl}| e^{i\alpha_{hkl}} e^{-2\pi i(hx+ky+lz)} \tag{1.15}$$

By using $\alpha = 2\pi\alpha'$

$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l |F_{hkl}| e^{2\pi i\alpha'_{hkl}} e^{-2\pi i(hx+ky+lz)} = \frac{1}{V}\sum_h\sum_k\sum_l |F_{hkl}| e^{-2\pi i(hx+ky+lz-\alpha'_{hkl})} \tag{1.16}$$

The structure factor for the reflection F_{hkl} can be written as:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j+ky_j+lz_j)} \tag{1.17}$$

where f_j is the scattering factor and (x_j, y_j, z_j) is the fractional coordinates of the atom j in the unit-cell, which establishes the phase of its contribution.

1.7 PHASE PROBLEM

From the above equation, $\rho(x, y, z)$ is obtained from a Fourier series from structure factors. Structure factors describe the diffracted rays that produce the measured reflections. The full description of a diffracted ray includes three parameters: amplitude, frequency and phase. The amplitude $|F_{hkl}|$ is proportional to the square root of I_{hkl}, but the phase of F_{hkl} is not directly obtainable from experimental measurement. To compute $\rho(x, y, z)$, we should first solve the phase problem, which is the most demanding element of X-ray crystallography.

1.8 METHODS TO SOLVE THE PHASE PROBLEM

Let us restrict ourselves to macromolecular crystallography. Four methods are used to solve the phase problem. There are: heavy-atom method (or isomorphous replacement method), anomalous scattering method (also called anomalous dispersion), molecular replacement method and direct method. All these methods only yield estimates of phases, which must be improved before an interpretable electron density map can be obtained. Initially, in these methods, phases are estimated for a limited number of reflections, and subsequently, phases determination must be extended to include as many reflections as possible.

1.8.1 The heavy-atom method (isomorphous replacement)

1.8.1.1 The Patterson function

In 1935, Patterson published a classical paper, which pointed out that Fourier

calculation could be carried out by using the phaseless quantity $|F|^2$ as coefficients. He showed that whereas the usual synthesis with F's as coefficients showed the distribution of atoms in the cell, the map calculated with $|F|^2$ gave peaks corresponding to all of the interatomic vectors. Thus a peak at a point uvw in a Patterson map indicates that atoms exist in the crystal at x_1, y_1, z_1 and x_2, y_2, z_2, such that

$u = x_1 - x_2 \qquad v = y_1 - y_2 \qquad w = z_1 - z_2$

For a molecule containing N atoms in a unit-cell, the Patterson synthesis will show N^2 peaks, corresponding to the N possible vectors that can be drawn from each of the N atoms. As the great size of origin peaks does not provide useful structural information and may cause trouble in scaling the computer output, it is often eliminated during the sharpening operation.

1.8.1.2 Patterson symmetry

Although the atomic positions in a real crystal can be distributed in any of the 230 space groups, vector peaks in the corresponding Patterson map are limited to only 24 space groups. This simplification reflects the loss of information suffered when proceeding from the phased F's to the phaseless $|F|^2$'s. The symmetry of the Patterson group is often higher than that of the crystal itself.

All Patterson maps will be centrosymmetric, regardless of the space group of the atomic distribution from which it is derived. Their lattice type is the lattice type of the original space group (P, C, F etc.). The Patterson space group is derived from the original crystal space group by replacing all translational symmetry elements (screw, glides) by their corresponding non-translational elements (axes, mirrors) and by adding

a center of symmetry, if it is not already present.

The symmetry elements of the crystal space group do not necessarily appear as such in the Patterson map. As aptly pointed and named after the discoverer, Harker lines and planes correspond to the usually high average intensity of certain regions of the reciprocal space group information in the Patterson map. They arise because the vectors between corresponding atoms of molecules related by symmetry elements have one or two constant coordinates. With the corresponding vectors positions in the Harker sections, we will be able to calculate the related atomic positions in the crystal (Buerger, 1959).

1.8.1.3 Heavy-atom derivative preparation

The heavier an atom the easier it is to locate by means of the Patterson map and the more it tends to determine the phases and intensities of all the reflections. In other words, the heavy-atom provides a better phasing model. However, if the atom is too heavy compared to other atoms in the structure, its dominance becomes too great and the comparison of $|F_o|$ and $|F_c|$ becomes relatively insensitive to the positions of the light atoms. Thus the uncertainty in the light atom coordinates increases. The reliability of bond length measurements decreases and in extreme cases, the light atoms may not be found at all.

A convenient rule that has been used as a guide in the selection of heavy-atom is

$$\sum Z_{heavy}^2 / \sum Z_{light}^2 \approx 1 \tag{1.18}$$

Note that it is better to err on the side of too light an atom instead of a too heavy-atom

(Stout, 1989).

In the most common technique, protein crystals are soaked in solutions of heavy ions. Several diffraction criteria define a promising heavy-atom derivative. First, the derivative crystals must be isomorphic with native crystals, that is, the incorporation of a heavy-atom must not disturb the crystal packing or conformation of the protein and unit-cell dimensions are not changed too much. The second criterion is that there must be measurable changes in a modest number of reflection intensities. Finally, the derivative crystal must diffract to a reasonably high resolution (about 3.0 Å), although it is not necessary to be as high as that of the native data. The phase extension step can produce phases for higher angle reflections from good phases of reflections at lower angles. Normally, at least two, and often more derivatives are required.

1.8.1.4 Heavy-atom determination

The most powerful tool used in determinating the heavy-atom coordinates is the Patterson function P(u,v,w), a variation of the Fourier series to compute ρ(x,y,z) from structure factors. The coordinates (u,v,w) locate a point in a Patterson map, in the same way that coordinates (x,y,z) locate a point in an electron density map. The coefficient of each term is the square of one structure amplitude, which is proportional to the measured reflection intensity. Thus the Patterson function in general form is:

$$P(u,v,w) = \frac{1}{V}\sum\sum\sum |F_{hkl}^2| \cdot e^{-2\pi i(hu+kv+lw)} \tag{1.19}$$

A Patterson map, which is a contour map of P(u,v,w), displays peaks at locations

corresponding to vectors between atoms. So there are more vectors between atoms than there are atoms, that is, a Patterson map is more complicated than an electron density map. But if the structure is simple, with only a few heavy-atoms in the unit-cell, the Patterson map may be simple enough to locate the atoms. Thus a difference Patterson map, instead of a Patterson map, is used to eliminate the enormous number of peaks representing vectors between light atoms in the protein.

To obtain the Patterson function solely for the heavy-atoms in derivative crystals, a difference Patterson function is constructed. Its coefficients $(\Delta F)^2 = (|F_{PH}| - |F_P|)^2$, the difference between the structure factor amplitudes with and without the heavy-atom reflects the contribution of the heavy-atom alone. The difference Patterson function is

$$\Delta P(u,v,w) = \frac{1}{V}\sum_h\sum_k\sum_l \Delta F_{hkl}^2 \cdot e^{-2\pi i(hu+kv+lw)} \tag{1.20}$$

If heavy-atoms bind to the protein at equivalent positions, heavy-atom peaks in the Patterson map can be found on the Harker sections. Based on this, the heavy-atoms can be located in the unit-cell. The structure factor with both amplitude and phases can be calculated with equation (1.17).

1.8.1.5 Protein phase determination

Compare a single reflection of amplitude $|F_P|$ in native data and the corresponding reflection of amplitude $|F_{PH}|$ from a heavy-atom derivative. The difference in the amplitudes ($|F_{PH}| - |F_P|$) is the contribution of the heavy-atom.

By vector algebra,

$$F_{PH} = F_H + F_P \tag{1.21}$$

For each reflection, we wish to know F_P (remember that if we can get the phase information of F_P, we will be able to calculate ρ) using

$$F_P = F_{PH} - F_H \tag{1.22}$$

We know $|F_{PH}|$ and $|F_P|$ from measuring the reflection intensities I_{PH} and $I_{P,}$ but not their phase angles. We will be able to know F_H, including its phase angle, after we locate the heavy-atom position and calculate its structure factor. Place the vector $-F_H$ at the origin of the Argand diagram and draw a circle of radius $|F_{PH}|$, centered at the head of the vector $-F_H$, Figure 1-5. All points on this circle will equal $|F_{PH}|$, that is the head of F_{PH} lies somewhere on this circle of radius $|F_{PH}|$. Now, add a circle of radius $|F_P|$ centered at the origin. We know that the head of the vector F_P lies somewhere on the small circle. Equations 1.21 and 1.22 above hold good only at points where the two circles intersect. Thus the phase angles of the two vectors F_P^a and F_P^b that terminate at the points of intersection of the circles are the only possible phase angles for the reflection. One heavy-atom derivative allows us to determine two values for α_{hkl} for each reflection hkl. To get the correct solution, a second derivative is used, which will agree better with one of the two solutions from the first derivative. In order to resolve the phase ambiguity from the first heavy-atom derivative, the second heavy-atom must bind at a different site on the protein from the first. If two heavy-atoms bind at the same site, the phase of F_H will be the same, because the phase of an atomic structure factor depends only on the location of the atom in the unit-cell and not its identity. In

practice, it often takes three or more heavy-atom derivatives to produce enough phase estimates. Obtaining phases with two or more derivatives is called the method of multiple isomorphous replacement (MIR). This is the method by which most protein structure is determined. When promising phases are available, Fourier synthesis is carried out to calculate $\rho(x, y, z)$.

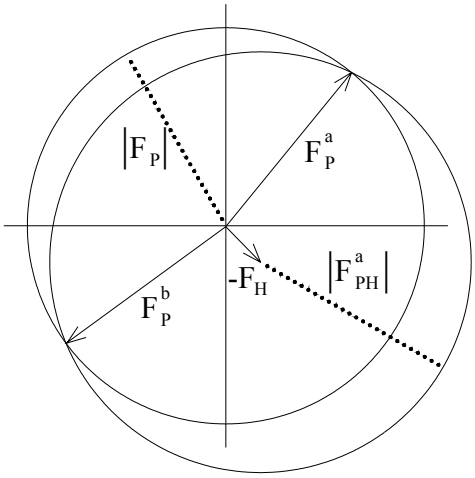


Figure 1-4. The principle of isomorphous replacement

1.8.2 The MAD method

1.8.2.1 Anomalous scattering

Recall elements absorb X-rays as well as emit them. The absorption drops sharply at wavelengths just below their characteristic emission wavelengths, called an absorption edge. This sudden change in absorption as a function of λ is called anomalous scattering. An element exhibits anomalous scattering when the X-ray wavelength is near the element's absorption edge. As a result of this absorption, Friedel's law does not hold (i.e., the reflections *hkl* and *-h-k-l* will not have equal intensity).

Absorption edges for light atoms in the unit-cell are not near the wavelength of X-rays used in crystallography. So carbon, nitrogen and oxygen do not contribute to anomalous scattering. However, absorption edges of heavy-atoms are in the experimental wavelength range. At the synchrotron, as the wavelength of X-rays is tunable, X-ray data can be collected under conditions that maximize the anomalous scattering by the heavy-atom.

1.8.2.2 Extracting phase from anomalous scattering

When the X-ray wavelength is near the heavy-atom absorption edge, a fraction of the radiation is absorbed by the heavy-atom and reemitted at an altered phase angle. At this wavelength, $F_{hkl} \neq F_{-h-k-l}$.

The effect of anomalous scattering on a given structure factor F_{PH} in the heavy-atom data is depicted in vector diagrams, Figure 1-6, as consisting of two perpendicular contributions, the real ΔF_r and the imaginary ΔF_i.

When data are collected at the wavelength λ_1, which is away from the absorption edge, there is no anomalous scattering. We use $F_{PH}^{\lambda_1}$ to represent the structure factor. When data are collected at the wavelength λ_2 near the absorption edge of the heavy-atom, anomalous scattering happens. We use $F_{PH}^{\lambda_2}$ to represent the structure factor:

$$F_{PH}^{\lambda_2} = F_{PH}^{\lambda_1} + \Delta F_r + \Delta F_i \tag{1.23}$$

At the wavelength λ_1, Friedel's law is still good, $|F_{hkl}| = |F_{-h-k-l}|$, and $\alpha_{hkl} = -\alpha_{-h-k-l}$. So

$F_{PH}^{\lambda_1-}$ is the mirror image of $F_{PH}^{\lambda_1+}$ in the real axis. The real contribution ΔF_r^+ and ΔF_r^- to the reflections of a Friedel pair are, like the structure factors themselves, reflections of each other in the real axis. But, the imaginary contribution to ΔF_i^+ is the inverted reflection of ΔF_i^-. That is, ΔF_i^- is obtained by reflecting ΔF_i^+ in the real axis and then reversing its sign or pointing it in the opposite direction. Because of this difference between the imaginary contribution to the reflections, $F_{PH}^{\lambda_2-}$ is not the mirror image of $F_{PH}^{\lambda_2+}$. From this disparity between Friedel pairs, the phase information can be extracted.

The magnitude of anomalous scattering contributions of ΔF_r and ΔF_i for a given element are constant (which can be obtained from the International Tables of crystallography) and roughly independent of the reflection angle θ. The phases of ΔF_r and ΔF_i depend only on the position of the heavy-atom in the unit-cell. If the heavy-atom is located by Patterson methods, the phases can be computed. Similarly, the $F_{PH}^{\lambda_1}$ can be solved by a vector diagram, where

$$F_{PH}^{\lambda_1+} = F_{PH}^{\lambda_2+} - \Delta F_r^+ - \Delta F_i^+ \tag{1.24}$$

To solve this equation, draw the vector $-\Delta F_r^+$ with its tail at the origin and draw $-\Delta F_i^+$ with its tail on the head of $-\Delta F_r^+$. With the head of $-\Delta F_i^+$ as the center, draw a circle of radius $\left|F_{PH}^{\lambda_2+}\right|$, which is the amplitude of the reflection in anomalous scattering. The head of the vector lies somewhere on the circle. If the phase is known, the head is located. Draw a circle of radius $\left|F_{PH}^{\lambda_1+}\right|$ with its center at the origin, representing the structure factor amplitude of the same reflection in the non-anomalous

scattering data set. The intersecting points will establish the phase of this reflection as either F_a or F_b that can meet the vector equation. But with the diagram, we can not tell which of the two phases is correct.

Remember the Friedel partner of the reflection $F_{PH}^{\lambda_2+}$ is not the mirror image of $F_{PH}^{\lambda_1-}$. This information will help us locate the right phase. As mentioned above, ΔF_r^- equals ΔF_r^+, $F_{PH}^{\lambda 1-}$ equals $F_{PH}^{\lambda_1+}$ and ΔF_i^- equals $-\Delta F_i^+$. So, the above equation (1.24), if replaced by $F_{PH}^{\lambda_2-}$, will be

$$F_{PH}^{\lambda_1+} = | F_{PH}^{\lambda_2-} | - \Delta F_r^+ + \Delta F_i^+ \tag{1.25}$$

Another vector diagram can be constructed as discussed next. Place the vectors $-\Delta F_r^+$ and ΔF_i^+ head to tail at the origin, and draw a circle of radius $| F_{PH}^{\lambda_2-} |$ centered at the head of $+\Delta F_i^+$. Finally, draw a circle of radius $F_{PH}^{\lambda_1+}$ centered at the origin. The circles intersect at two points giving two solutions of the equation. One of the solutions will correspond to F_a in the previous equation, which is the right phase of $| F_{PH}^{\lambda_1+} |$. Once the phase of F_{PH} is obtained, and F_H is known from calculating the heavy-atom structure factors after locating the heavy-atom by Patterson methods, the vector F_P is simply the vector difference of F_{PH}-F_H. The phase of this reflection in the native data can also be defined.

Currently, MAD is the fastest growing method of structure determination in macromolecular crystallography. As different wavelength data are collected on a single crystal, non-isomorphism is no longer a problem and more accurate estimates of phase can be obtained. Also with new generation of synchrotron radiation, high

resolution and accurate data enable enable MAD to work really well and quickly. As it is necessary to collect the data with powerful synchrotron source x-ray, the greatest impediment of MAD today is access to synchrotron facility for the data collection.

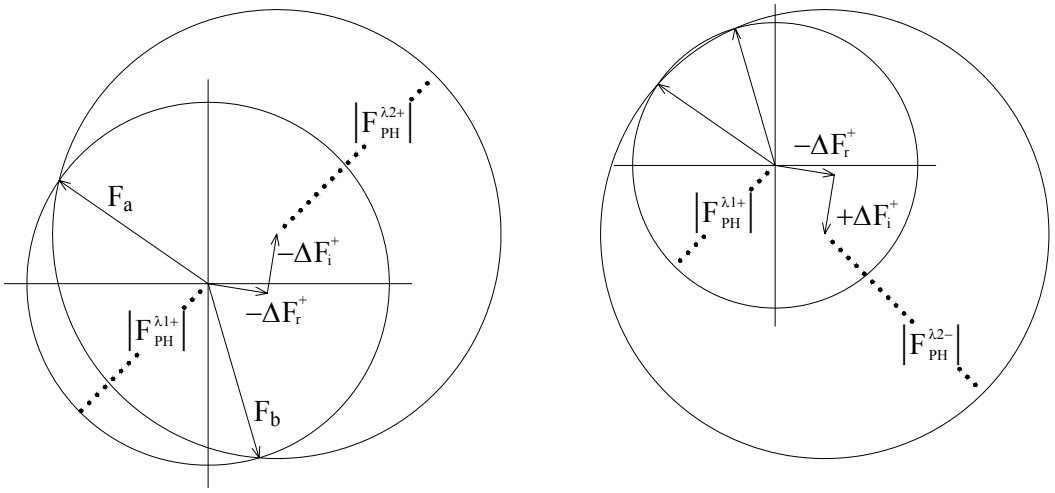


Figure 1-5. The principle of anomalous method

1.8.3 Direct methods

The above methods involving heavy-atoms apply to large molecules with 500 or more non-hydrogen atoms. For small molecules (up to 200 atoms), phases of reflections can be determined by *ab initio* direct methods, which rely on the existence of mathematical relationships among certain combinations of phases. With direct methods, all that is required is a single set of native diffraction intensities. The direct methods use sophisticated probability theory and assumption of approximately equal and resolved atoms to estimate reflection phase from measured intensities (Uson, 1999).

The crucial break through of direct method is the introduction of dual space

iteration, known as 'Shake-and-Bake' (Miller, 1993). Here, starting with random atoms with constraints such as the type and number of atoms, consistency with the Patterson function and interatomic distance restrictions and so on, phases are calculated from the starting atoms and different strategies are applied in the reciprocal space to try to improve these phases. Minimization of the minimal function R(ϕ) and the tangent formula (tan ϕH) are the two most frequently applied strategies.

The minimal function is the weighted mean square difference between the current and statistically expected cosines of the sums of three phases. For a given set of structure factors with random or non-random phases, R(ϕ) is minimized by a parameter shift method; each phase may be shifted once or twice by a given value and the shifted phases are adopted if they lead to reduction in the minimal function. Next, a map is generated and from the map, atoms are picked. These atoms are then used to calculate new structure factors, the phases of which are used as the starting point for a new round of minimizing R. The procedure is iterated until convergence. Many runs, each using a different starting point in the reciprocal space, are required in order to find the right solution. This is the default algorithm in the SnB program (Abrahams, 1998).

The tangent formula forms the basis of most conventional direct method programs, either to refine all the phases or to derive phases for the remaining reflections from the phases that are best determined by the current atoms. This is default in the SHELXD program.

So far, the direct method has successfully solved the structure of triclinic lysozyme with 1000 atoms, which contains no atom heavier than sulphur (Deacon, 1998). With heavier elements, larger structures (for example, cytochrome c_3 with 2024

atoms) can be solved (Frazão, 1998). The resolution limit seems to be a tougher barrier and structures tend to diffract to lower resolution as they get bigger. Most structures solved by the direct method had a resolution of 1.1 Å or better. However, this limit can be relaxed if heavy-atoms are present. It can be expected that structures with more atoms will become amenable to direct methods as computers become more powerful.

1.8.4 Molecular replacement: related proteins as phasing models

If the structure of a new protein is similar to that of a known protein, the known protein can be used as a phasing model to solve the phase problem of the new protein without making any heavy-atom derivatives. The method that entails calculating initial phases by placing a known protein model in the unit-cell of the new protein is called the molecular replacement method.

1.8.4.1 Isomorphous phasing models

If the phasing model and the new protein are isomorphous, for example, small ligand is soaked into a protein crystal, the phase from the ligand free protein structure can be used directly to compute $\rho(x,y,z)$ for the ligand bound protein crystal (using the data of this crystal).

$$\rho(x,y,z)=\frac{1}{v}\sum_{h}\sum_{k}\sum_{l}|F_{hkl}^{new}|e^{-2\pi i(hx+ky+lz-\alpha_{hkl}^{'mod\,el})} \tag{1.26}$$

Here amplitude $\left|F_{hkl}^{new}\right|$ are obtained from the native intensities of the new protein and the phases $\alpha^{'mod\,el}$ are those of the phasing model. After interactive steps of phase improvement, the protein is well placed in the unit-cell and by synthesizing a

difference Fourier map, the ligand is easily located in the unit-cell.

1.8.4.2 Non-isomorphous phasing models

If the phasing model is not isomorphous with the desired structure, more work needs to be done. In order to use known protein as a phasing model, the structure of the model should first be oriented to the structure of the new protein in its unit-cell and phases are calculated for the properly oriented model. That is, the best and right orientation of the model in the unit-cell of the target protein must be found.

The Patterson map from the known structure reveals that the final map is independent of the position of the structure in the unit-cell, but only depends on the orientation of the structure in the unit-cell. This observation suggests that the Patterson map can be used to determine the best orientation of the model in the unit-cell of the new protein. That is if the model and the new protein are indeed similar, their same orientation will give very similar Patterson maps. So Patterson maps for various model orientations are computed and compared with the Patterson map of the target protein. By this way, the best orientation of the model can be found, and based on this orientation, the best position of the models will be searched. For orientation search, the parameter called model Patterson function $P^{Model}(u,v,w)$ is checked to monitor the search. A value of zero for P^{Model} means the trial orientation has no peak fit in the target protein Patterson map. A high value means that the trial orientation has peaks at all locations of peaks in the Patterson map of the target protein, which means this trial might be the correct orientation. For position search, the structure factor (F_{calc}) of the model is calculated and the amplitude is compared with the measured amplitude $|F_{obs}|$ obtained from different intensities of the new proteins. This criterion can be expressed

as R factor.

$$R = \frac{\sum || F_{obs} | - | F_{calc} ||}{\sum | F_{obs} |} \tag{1.27}$$

If the observed and calculated amplitude agrees with each other, R is small. For a protein, an R value of 0.3 ~ 0.4 for the best placement of phasing model have often provided adequate initial estimates of phases.

1.9 IMPROVEMENT OF ELECTRON DENSITY MAP AND MODEL BUILDING

After getting the phase information from one of the above four methods, the electron density $\rho(x,y,z)$ can be calculated using equation 1.16. Because the initial phases are roughly estimated, the first map may be uninformative and uninterpretable. Crystallographers improve the map by an interactive process.

1.9.1 Weighting factor

For each set of phases, a weighting factor that decides the precision of the phase is selected. The value of the weighting factor W is between 0 and 1. A Fourier term containing a phasing estimate of low reliability will be multiplied by a low weighting factor in the Fourier series computation of $\rho(x,y,z)$. This reduces the bias from a reflection whose phase is poorly estimated. Conversely, a term containing a phase of high reliability will be given full weight (weighting factor of 1.0) in the series:

$$\rho(x, y, z) = \frac{1}{v} \sum_h \sum_k \sum_l W_{hkl} | F_{obs} | e^{-2\pi i(hx+ky+lz-\alpha^{'calc})} \tag{1.28}$$

1.9.2 Improving the map

1.9.2.1 Solvent flattening

The first map of acceptable quality will display the protein that is barely distinguishable from bulk water. The process that assigns the regions of a protein molecules a high value of $\rho(x,y,z)$ and a low value of ρ for all surrounding areas of bulk solvent is called solvent flattening. In practice, the unit-cell is divided into a grid of regularly spaced points. At each point, the value of $\rho(x,y,z)$ in the F_o map is evaluated. At each grid point, if ρ is negative, it is reassigned a value of zero; if ρ is positive, it is assigned a value equal to the average value of ρ within a defined distance of the grid point. By this manner, the electron density map will be smoothened. Small, random fluctuations in density will be eliminated, and the map is essentially divided into two types of region, relatively high density (protein) and relatively low density (solvent). The overall amplitude of the map is increased until the ratio of high density to low density agrees with the ratio of protein to solvent in the crystal. The contrived function $\rho(x,y,z)$ is now used to calculate the structure factor with new phases. Along with $|F_{obs}|$'s derived from original measured intensities, $\rho(x,y,z)$ is calculated again. If the new phase estimates are better, the new electron density map will reveal better details by clearly showing the protein-solvent boundary.

1.9.2.2 Phase extension

Generally, in the heavy-atom methods described above, data in the intermediate resolution range (up to 3.0 Å) are used. We will be able to gradually increase the number of terms in the Fourier series by adding native intensities with higher

resolution than the current model. This process is called phase extension. This work must be done gradually and judiciously to avoid introducing any bias. In this manner, low resolution phase is improved and phase assignments are extended to a higher resolution.

1.9.2.3 Non-crystallographic symmetry averaging

Non-crystallographic symmetry is symmetry that exists locally within the asymmetric unit of the crystal. By knowing the symmetry elements, crystallographers can produce averaged maps in which noise will tend to cancel out and can be used as phase restriction to improve phasing. This method is especially helpful in a system of a high degree of symmetry.

1.9.3 Model building

A molecular model must fit realistically with the calculated electron density map. A realistic three-dimensional display of the electron density map will allow the user to construct and manipulate molecular models to fit the map. Amino acids properties, stereochemistry, geometrical constraints of the peptide bond, the dihedral angles of the main chain etc. must be well realized during model building. Relatively, side chains are less restricted than the main chain. Side chains have preferred rotomer positions. During the initial steps of model building, it will be very helpful if large secondary structural elements are recognized. Once the secondary structures are known, many of the breaks in the density can be confidently assigned based on the constraints of secondary structures. Helices are mostly obvious in 5.5 Å maps while β-sheets tend to break up at resolution less than 3.0Å. Another milestone in structure determination is finding the first match of sequence in the map. As chain termini are often ill-defined,

we need a foot-hold for alignment of the sequence where the map is sharp. Often, the key is a short stretch of sequence containing several bulky hydrophobic residues. Because they are hydrophobic, they are likely to be in the interior where the map is clear. From such a foot-hold, detailed model building can begin. Regions that can not be aligned with the sequence are often built with polyalanine. In this manner, we build as many atoms of the model as possible into the electron density. Using this fitted model, new phase are calculated. If the model is correct, the subsequent map calculated with the new phase will show additional feature, leading to further improvement of the model while wrong model will degrade the map. Iteratively, we follow the good sections of the map and add more protein segments and complete the model (McRee, 1999a).

1.9.4 Refinement

When the structure nears completion, structure refinement is performed to improve the agreement of the model with the original intensity data.

1.9.4.1 Least-squares methods

Commonly, the refinement process consists of a version of least-squares fitting. In a simple least-squares method, the aim is to find a function $y = f(x)$ that fits a series of observations (x_1, y_1), (x_2, y_2).....(x_i, y_i). The solution to the problem is a function $f(x)$ for which the sum of the squares of distances between the data points and the function, itself is as small as possible.

$$D = \sum_i w_i (y_i - f(x_i))^2 \tag{1.29}$$

That is, *f(x)* is the function that minimizes *D*, the sum of the squared difference between observed (y_i) and calculated $f(x_i)$ values.

The differences are squared to make them all positive and prevent a large number of random difference add up to drive *D* toward 0. W_i is an optional weighting factor that reflects the reliability of the observation. Normally in statistics, W_i is the deviation computed from multiple measurements of the same data point (x_i, y_i).

1.9.4.2 Crystallographic refinement

In the case of crystallography, with the model built from the electron density map, we can calculate the structure factor ($|F_c|$). With all the atoms *j*, by comparing the calculated structure factor ($|F_c|$) and the observed structure factor $|F_o|$, we will be able to judge whether the model atoms are at right positions. In the least-squares terminology, we want to select the atom positions that minimize the square difference between corresponding $|F_c|$'s and $|F_o|$'s.

$$\phi = \sum_{hkl} W_{hkl} (|F_o| - |F_c|)^2_{hkl} \tag{1.30}$$

ϕ is the sum of the squares of differences between observed and calculated amplitudes. The sum is taken over all reflections hkl in use. Each difference is weighted by the term $w_{hkl,}$ a number that depends on the reliability of the corresponding measured intensity.

1.9.4.3 Molecular dynamics refinement

This approach treats the model as if its energy decreases when its fit to the

native crystallographic data improves. In refinement by simulated annealing, the model is allowed to move to a high temperature and lifted out of local energy minima. Then, the model is cooled slowly to find its preferred conformation at the temperature of diffraction data. Throughout the process, the computer is searching for the conformation of the lowest energy with the assigned energy is partially dependent on agreement with diffraction data.

1.9.4.4 Additional parameters for refinement

Besides the atom positions that are included in refinement, other parameters are included as well. One is the temperature factor B_j of an atom j, a measure of how much the atom oscillates around the position specified in the model. As atoms are not static in the model, they always vibrate around their positions, especially atoms at side chain termini. Diffraction is affected by this variation in atomic positions, so it is realistic to assign a temperature factor to each atom.

Another parameter included in refinement is the occupancy n_j of each atom j, a measure of the fraction of the atom j that actually occupies the position specified in the model. For example, if two conformations occur with equal frequencies, then the involved atoms receive occupancies of 0.5 each for their two possible positions. By including occupancies among the refinement parameters, estimates of the frequency of alternative conformations is obtained, which gives some additional information about the dynamics of protein molecules.

A least-squares procedure will find the minimum that is nearest the starting point, so it is important that the starting model parameters are near the global minimum. Otherwise, the refinement will converge into an incorrect local minimum,

from which it can not extract itself. The greatest distance from the global minimum from which refinement will converge properly is called the radius of convergence. Additional constraints and restrains used in the refinement computation will increase the radius of convergence and thus increase the probability of finding the global minimum. A constraint is a fixed value for certain parameter. We may constrain all occupancies to a value to 1.0. A restraint is a subsidiary condition imposed on the parameters, such as the condition that all bond lengths and bond angles are within a specified range of values. With the help of modern computers, least-squares restraints, including bond length, bond angle and conformational energies and energies of non-covalent interactions are added to the overall energy refinement to find the structure of the lowest energy.

1.10 FINAL STRUCTURE

After several rounds of refinement and map fitting, the model is slowly converged to the final model. The program for iterative real and reciprocal space refinement is monitored by computing the difference between the measured structure factor $|F_o|$ and the calculated structure factor $|F_c|$ from the current model. When the model converges to the correct model, the difference between measured *F*'s and calculated *F*'s will also converge

$$R = \frac{\sum || F_{obs} | - | F_{calc} ||}{\sum | F_{obs} |} \tag{1.31}$$

A desirable target *R* factor for a protein model should be less than 0.3. Occasionally, small and well ordered protein structure may refine to about R = 0.1

In addition to monitoring *R* as a convergence indicator, various structural parameters (like the model is chemically, stereochemically and conformationally reasonable) are important. Above geometry is normally checked with program Procheck (Laskowski, 1993). In a well refined model, the root mean square (RMS) deviation of bond lengths should be no more than 0.02 Å and bond angles is less than 4°. Also, there should be no D-amino acid, peptide planes are nearly planar, the back bone conformational angles ϕ and ψ fall in the allowed regions and torsional angles in side chains lie within a few degrees of stable and staggered conformation. Then the final model is deposited into Protein Data Bank with AutoDep Input Tool (*http://pdbdep.protein.osaka-u.ac.jp/adit/*).

CHAPTER 2 HEX1 CRYSTAL LATTICE IS REQUIRED FOR WORONIN BODY FUNCTION IN *NEUROSPORA CRASSA*

2.1 INTRODUCTION

2.1.1 Discovery of Woronin body and its function

In 1864 Woronin, a Russian mycologist, reported the characteristics of a distinct type of organelle in the fungus *Ascobolus pulceherrimus* (Woronin, 1864). This organelle is described as a dense-core vesicle and is generally seen in the vicinity of the septal pore, which connects cellular compartments and allows transportation of the cytoplasm and subcellular organellels between cells. In the subsequent 135 years, this organelle, named as Woronin body by Buller (1933), has been identified in more than 50 fungal species. However, their cellular origins and biochemical composition have remained obscure. They appear to be restricted to the filamentous Ascomycotina, e.g., *Neurospora crassa* and *Aspergillus nidulans,* and the Deuteromycotina, filamentous fungi in which a sexual stage has not been identified (Markham, 1994; Markham and Collinge, 1987). Filamentous Ascomycotina form vast networks of hyphae in which the cytoplasm, including large organelles such as nuclei, can flow from compartment to compartment, moving though small pores in the septa that separate compartments. This type of cellular organization will enable the extremely rapid rate of tip extension the filamentous fungi can achieve. However, this mode of

cellular organization also carries risks. When hyphae are damaged, the cytoplasm will bleed out rapidly. The Woronin body plays an important role in this emergent situation. They will detach from the cell membrane and plug the septal pores within a few seconds with the help of chitin, see figure 2-1. (Collinge and Markham, 1985). Subsequently, it initiates a sequence of events that ultimately leads to the resealing of the plasma membrane and the resumption of polarized growth. This mechanism is essential to prevent the loss of the cytoplasm as fungi growing in nature are subject to constant physical damage from animals, wind, etc. (Jedd and Chua, 2000; Tenney et al., 2000).

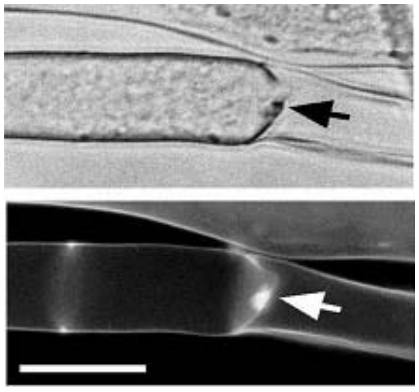


Figure 2-1. A Woronin body occluded the septal pore is surroundedby chitin soon after wounding of N. crassa mycelia. Arrow indiactes a Woronin body that has plugged the septal pore. Upper panel, phase contrast; lower panel, chitin staining.Scale bar = 10 μm. (Jedd and Chua, 2000)

2.1.2 The category of Woronin body

Most early information concerning the nature of the Woronin body came from microscopic analysis. As the size of the Woronin body is close to the limit of resolution of the light mocroscope, it is described as refractile bodies near the septal pore. They fall into two structural classes, spheres and hexagonal rods. Early research suggests that the hexagonal Woronin body is an ergosterol crystal. But it turned out to be incorrect, as specimens of the Woronin body were impervious to solvents that dissolved sterols, but disappeared when treated with proteases, which suggests that the core of the Woronin body is actually proteinaceous (Hoch and Maxwell, 1974). Furthermore, the typical hexagonal Woronin body is identified in ergosterol synthesis defective mutants in *Neurospora crassa* (Armentrout and Maxwell, 1974). In electron micrographs, they are typically identified as 150 to 500 nm organelles, bound by a single membrane. Normally, three to six copies of the Woronin body are present in each cell (Tenney et al., 2000). The Woronin body of *N. crassa* belongs to the hexagonal rod class with a size of 400 x 700 nm, which is slightly larger than the septal pore (Trinci and Collinge, 1974).

2.1.3 Hex1 is responsible for the function of Woronin body

There has not been much progress on the biochemical and biosynthesis study of the Woronin body until the molecular component of this unusual organelle has been identified. In 2000, Jedd and Chua isolated a fraction enriched with the Woronin body from *N. crassa* by sucrose cushion centrifugation and a high salt wash. N-terminal sequence analysis of an abundant 19 kDa protein showed that it was encoded by a gene (GenBank Accession No. AF001033). Jedd and Chua named this gene *Hex1*. Anti-Hex1 antibodies bound uniformly and exclusively to the matrix of the *N.*

crassa Woronin body. Deletion of the *Hex1* gene by homologous recombination in *N. crassa* caused the disappearance of the Woronin body. The Hex1 deletion strain was capable of producing aerial hyphae and conidia, which shows that Hex1 does not have any important role in either vegetative growth or development. But this strain failed to seal the septal pore and was defective in the reinitiation of polarized growth, which the wild type does when syncytia are severed. Furthermore, the mutant strain extruded large puddles of the cytoplasm through the wounded proximal septal pore. These data indicate that Hex1 is the major component of the Woronin body and is directly responsible for the function of the Woronin body (Jedd and Chua, 2000; Tenney et al., 2000).

2.1.4 Woronin body is a new type of peroxisome

Because of the size, electron density, and crystalline inclusion, people suggested that the Woronin body might be a specialized type of microbody (Wergin, 1973), while the isolated Woronin body did not show significant catalase activity (Head et al., 1989). Latest work by Jedd confirmed the peroxisomal origin of the Woronin body with the evidence of antibodies that were directed to the peroxisome targeting signal (PTS1) that is present in the *N. crassa* Woronin body.

Peroxisomes were discovered as biochemical entities by De Duve's group (Baudhuin, 1965). They were identified as small sedimentable particles containing marker enzymes that distinguished them from other known organelles. They are a ubiquitous class of subcellular organelles that carry out diverse functions associated with anabolic and catabolic pathways, including peroxide metabolism, the β-oxidation of fatty acids and the biosynthesis of phospholipids (Subramani, 1998). There are

likely to be at least two pathways for targeting proteins to the peroxisomal matrix, both of which require a specific peroxisome targeting signal (PTS) and a cognate receptor (Olsen, 1998; Subramani, 1998). PTS1 is a simple tripeptide carboxy-terminal signal (amino acid sequence S/A/C-K/R/H-L/M) that is sufficient to mediate the peroxisomal targeting of reporter proteins (Gould, 1989; Keller, 1991). A second signal, PTS2, is found at the amino-terminal ends of proteins, which has a similar function (Swinkels et al., 1991). After translation, PTS1- and PTS2-containing proteins are recognized by distinct cytosolic receptors that target them to the peroxisomal membrane, where the core of the translocation apparatus is located (Subramani, 1998; Tabak, 1999). Peroxisomes are tailored to execute functions specific to organisms and even tissues. In plants, for example, leaf peroxisomes are specialized to participate in photorespiratory reactions, whereas glyoxysomes are specialized to function in the β-oxidation of fatty acids (Olsen, 1998). Peroxisomal identity seems to be determined by the differential expression of matrix components, all of which use similar targeting signals (PTS1 or PTS2). Hence, in plants, glyoxysomes and leaf peroxisomes are interconverted, depending on the developmental and physiological cues that regulate the expression of their respective enzyme constituents. Expressed from the constitutive glyceraldehyde-3-phosphate dehydrogenase (*GPD*) promoter, a GFP–Hex1 fusion protein was localized in a punctate pattern, characteristic of peroxisomes. The deletion of Hex1 leucine 176 (L176 of its PTS1 target signal SRL) altered the C-terminal tri-peptide to the amino acid sequence GSR and caused the GFP–Hex1 construct to be present in the cytoplasm, as did the deletion of *PEX13*, which encodes the yeast PTS1-recognition-factor receptor. Thus, Hex1 PTS1 is a functional peroxisome targeting signal, which also confirms that the Woronin body is a special peroxisome.

2.1.5 Hex1 has the characteristics of self-assembly

The geometric structure of the Woronin body suggests it might be crystalline. Hex1 and the Hex1(ΔL176) mutant, which alters the PTS1 signal, were overexpressed in yeast cells and subsequently examined by thin section electron microscopy. Hex1-expressing cells produced protein granules that were enclosed by a unit membrane and presented hexagonal profiles similar to those of the *N. crassa* Woronin body. No similar structures were observed in strains transfected with either Hex1(ΔL176) or an empty plasmid. These results show that Hex1 can assemble into hexagonal rods in the matrix of the yeast peroxisome. Recombinant Hex1 was seeded onto glass slides at a concentration of 1 mg ml^{-1} and incubated overnight in a humid chamber at room temperature. Under these conditions, Hex1 readily produced hexagonal crystals of about 10 μm long that could be seen by light microscopy. The *N. crassa* Woronin body has a simple geometric structure that suggests a crystalline composition. Similar geometry of Hex1 crystals assembled *in vitro* also indicates that the Hex1 protein has the characteristics of self-assembly (Jedd and Chua, 2000).

2.2 HEX1 STRUCTURE DETERMINATION

2.2.1 Purpose of Hex1 structure determination

Both the Hex1 crystal and Woronin body have a hexagonal rod profile. As the Hex1 protein shows the characteristics of self-assembly, it automatically raises the following questions. Does Woronin body formation rely on Hex1 self-assembly? Do the Woronin body and Hex1 crystal adopt the same crystalline lattice? What would the crystalline lattice be like? To answer all the above questions and understand the

relationship between Hex1 assembly and the biological function of the Woronin body, we have undertaken the crystal structure determination of Hex1.

2.2.2 Experimental methods

2.2.2.1 Expression and purification of native Hex1

The full length *N. crassa Hex1* gene was subcloned into the pGEX-6p-1 vector (Pharmacia) between the *Eco*RI and *Ba*mH1 sites, and the resulting construct (GJP 2-186) was transformed into the *E. coli* strain BL21. One colony was used to inoculate 5 ml LB medium with 100 μg ml^{-1} Ampicillin and grown at 37 °C at 200 rpm overnight. Next morning, this overnight culture was added to 1 liter LB medium with 100 μg ml^{-1} Ampicillin and growth was continued at 37 °C, 200 rpm in a INFORS AG CH-4103 incubator (Botlmingen / Switzerland) until $OD_{600\ nm}$ was 0.5. Cells were induced with 0.2 mM IPTG for 3 hours. The cells were then harvested by centrifuging at 5,000 g for 20 minutes using a Sorvall RC 26 Plus centrifuge and Sorvall SLA 1500 superlite rotor. The supernatant was carefully removed and the cells were suspended in 10 ml lysis buffer A (see Appendix A). The suspension was then passed through a French press (SIM-AMINCO spectronic instruments) at 1,000 psi pressure to completely lyse the cells. The slurry was then centrifuged at 20,000 g for 30 minutes using a Sorvall S-300 rotor. The insoluble portion was pelleted down. The supernatant was carefully transferred to a Falcon tube and mixed with 2 ml glutathione sepharose resin (Amersham Pharmacia Biotech Inc.), which was prewashed with 40 ml lysis buffer A, and incubated at 4 ºC for 30 min to aid the protein bind to the resin. The resin was poured to a 20 ml Econo-pac disposable chromatography column (Bio-Rad). The column was washed with 40 ml wash buffer B1 four times and 40 ml wash buffer B2

twice. The resin was next washed with 2 ml cleavage buffer C twice and suspended in 2 ml cleavage buffer. Hex1 was released from the glutathione sepharose resin by cutting off the GST fusion with 80 units of the PreScission protease for each ml of the resin by incubating at 4 °C for a period of 12 hours. The protein was then concentrated to 3 mg ml^{-1}, frozen with liquid nitrogen and stored at -80 °C for later use. Generally, 1 liter culture gave about 2 mg pure Hex1 protein (Figure 2-2).

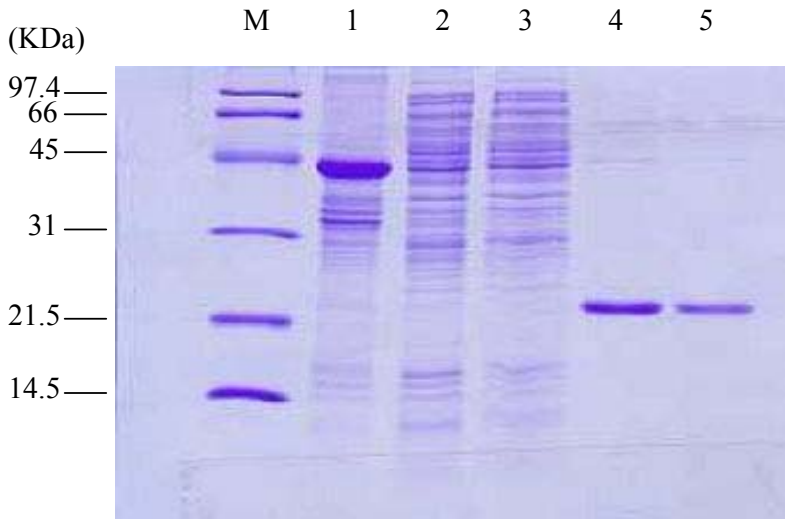


Figure 2-2. Purification of Hex1. Lane 1 is the total lysate. Lane 2 is the lysate that does not bind to the glutathione sepharose resin. Lane 3 is the wash elution of the resin. Lanes 4 and 5 are the final elution of Hex1 after cleaving the GST-fusion tag.

2.2.2.2 Selenomethionine Hex1 expression and purification

One colony from a fresh LB cell plate with 100 μg ml^{-1} Ampicillin was picked and inoculated to 5 ml LB and cultured at 37 °C, 200 rpm overnight. Next morning, 1 ml overnight culture was gently spinned at 1,300 g for 3 min on an Eppendorf 5145C

microfuge. The cells were resuspended in 1 ml of M9 medium and pelleted again. The pellet was resuspended in 50 ml pre-warmed M9 medium with 100 μg ml^{-1} Ampicillin. The culture was grown at 37 °C and 200 rpm until OD_{600} was about 0.8 (it normally took about 12 h). 50 ml of this cell culture was added to 1 L prewarmed M9 medium with 100 ug ml^{-1} Ampicillin and continued to grow to an OD_{600} of 0.6. The following amino acids were added, at the given concentrations, to the medium: Lysine, Phenylalanine and Threonine at 100 mg L^{-1}; Isoleucine, Leucine and Valine at 50 mg L^{-1}; L-selenomethionine at 60 mg L^{-1}. 15 min later, 400 μl of 500 mM IPTG was added to the medium to induce protein expression. The cell was continued to grow at 200 rpm and 37 °C for 6 hours and then harvested by centrifugation at 5,000 g at 4 °C for 5 min using a Sorvall RC 26 Plus centrifuge and Sorvall SLA 1500 superlite rotor.

The pellet was washed with 100 ml TDE buffer and then resuspended in 10 ml TDE buffer. 50μl of 500mM PMSF was then added to the suspension. The cells were lysed using a French press, as explained earlier. 1 ml of 10% Triton X-100 was added to 10 ml lysed slurry and mixed before incubating on ice for 10 minutes. Other protein purification steps were nearly identical to those used for the purification of the native protein, except TDET buffer was used as wash buffer. The protein was 95% pure, which was adequate for crystallization (Figure 2-3). The purified samples were analyzed by electrospray ionization mass spectrometry. The mass of Hex1 and Se-Hex1 were 19694.46 and 19951.76 Da, respectively. The mass difference clearly accounts that all five methionines of Hex1 had been replaced by selenomethionine (Coligan et al., 1995). The purified protein was concentrated to 3 mg ml^{-1} and frozen with liquid nitrogen and stored at -80 °C for later use.

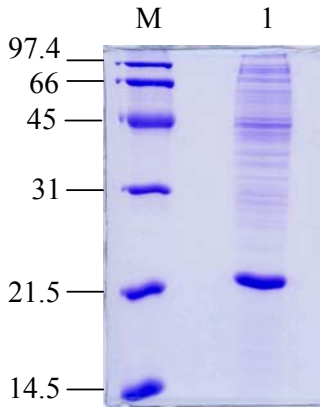


Figure 2-3. Purified selenomethionine Hex1, lane-1.

2.2.3 Hex1 crystallization

2.2.3.1 Native crystal

Native crystals were grown at 20 °C by the sitting drop vapour diffusion method. 4 µl of the Hex1 protein, at a concentration of 3 mg ml^{-1}, was mixed with an equal amount of reservoir buffer consisting of 60 mM Tris (pH 7.5), 210 mM NaCl and 30 mM $CaCl_2$. Small crystals were seen after 1 day. 5% (w/v) 2-methyl-2,4-pentanediol (MPD) was added to the reservoir 2 days later, which helped the crystal grow further to the size of 0.2 x 0.2 x 0.3 mm in one week. The shape of the crystal was generally a hexagonal prism, Figure 2-4.

Figure 2-4. Native Hex1 crystal

2.2.3.2 Selenomethionine crystal

Se-Met derivative crystals were grown by the same procedure. However, Se-Met crystals could not grow at the same condition that produced the native Hex1 crystal. Hence, the crystallization condition was re-screened with the Hampton screens I and II. Condition 34 of the Hampton screen I gave very tiny crystals. Careful optimization of the crystallization condition (by adjusting the salt content and the pH of the solution) finally yielded relatively bigger crystals at the condition of 0.1 M sodium acetate trihydrate and 2 M sodium formate (pH 7.0). The crystal shape was also a hexagonal prism. Compared to the native crystal, Se-Met crystals were smaller and they could only grow to 0.05 x 0.05 x 0.10 mm, Figure 2-5.

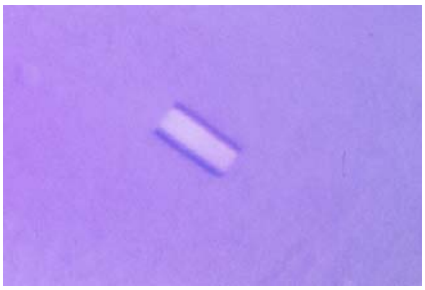


Figure 2-5. Selenomethionine Hex1 crystal

2.2.4 Hex1 data collection

Both the native and Se-met Hex1 crystals were flash frozen with a cryoprotectant containing 50% (v/v) mineral oil and 50% paratone (v/v). Diffraction data were collected at the beamline X12C, National Synchrotron Light Source, Brookhaven National Laboratory (BNL). All diffraction frames were processed with DENZO 1.96 and scaled using SCALEPACK (Otwinowski, 1997). The native crystal diffracted X-rays to 1.78 Å and the Se-Met crystal diffracted to 2.3 Å. The anomalous signal was obvious in the Se-Met crystal data, Table 2-1.

Table 2-1 Native and Selenomethionine crystal data

Unit-cell dimensions	a = b = 57.427, c = 196.975 Å, α=β=90°, γ=120°			
Space group	$P6_522$			
Data sets	Edge	Peak	Remote	Native
Wavelength (Å)	0.978707	0.978441	0.93001	0.97863
Resolution (Å)	50-2.3	50-2.3	50-2.3	99-1.8
Measured reflections	257,556	237,392	177,070	123,270
Unique reflections	15,990	15,981	16,348	18,047
Redundancy	11.9	11.0	7.7	4.3
Completeness, total/last shell	98.5/99.4	98.3/97.3	98.1/99.4	95.6/85.1
[1]R_{sym}, total/last shell	0.101/0.306	0.095/0.311	0.116/0.370	
0.055/0.159				

[1]$R_{sym} = \sum_{hkl}\sum_i[|I_i(hkl) - \langle I(hkl)\rangle| / I_i(hkl)]$.

2.2.5 Selenium position determination

The structure was determined with the scaled data of the Se-Met Hex1 crystal

collected at the remote, peak and inflection wavelengths (at the 3 Å resolution limit) using the SOLVE program (Terwilliger, 1999). All the five selenium positions were determined. The vectors generated from the heavy-atom positions according to the crystal symmetry just fit the peaks of the Harker section in the anomalous difference Patterson map generated by Xtalview (McRee, 1999b), which confirmed the solution. The statistics of the structure solution are listed in Table 2-2.

Table 2-2 SOLVE statistics for selenium determination

Criteria	MEAN	SD	VALUE	Z-SCORE
Pattersons	0.729	0.500	1.83	2.21
Cross-validation Fourier	7.40	9.68	95.3	9.08
NativeFourier CC x100	6.75	1.85	21.8	8.15
Mean figure of merit x100	-	6.49	55.4	8.53
Correction for Z-scores		-6.60		
Overall Z-score value		21.4		

The solution of the heavy-atom positions were also introduced to the program VECTORS (the CCP4 package, 1994), and the predicted vector positions correctly fit the anomalous difference Patterson map and confirmed that the heavy-atom positions were correct.

2.2.6 Electron density map

The five selenium positions were used in the PHASIT module (the PHASES package, Furey, 1997) for refinement. First, the phase angles of the reflections upto

3.0 Å were calculated. The phasing power of the heavy-atoms using the isomorphous component (Δf′) difference between the edge and remote data was 2.64 with a Figure of Merit of 0.451 for 6572 reflections. The phasing power of the heavy-atoms using the anomalous component (Δf″) difference between the peak and remote data was 4.28 with a Figure of Merit of 0.522 for 5371 reflections. The solvent content of this crystal was 47.24% and the Matthew's coefficient was 2.34 (Matthews, 1968). The phase angle values were improved by minimizing the featureless density of the solvent region. 50 cycles of solvent flattening was performed with the DOALL.SH script of the PHASES package. An electron density map was calculated using the solvent-flattened phases and improvement on the original MAD map was substantial. In the core region of protein molecules, some secondary structure elements were clearly recognizable. The connectivity of the density was good and chain tracing was readily possible. As the Se-Met Hex1 crystal actually diffracted to 2.3 Å, the phases were slowly extended to 2.3 Å resolution with the EXTND.SH script of the PHASES package. Electron density maps at 3.0, 2.5 and 2.3 Å resolutions were created with the GMAP.SH script of the same program.

The definitions of some of the parameters are given below.

Phasing power $= \dfrac{\sum \left| F_{H_{calc}} \right|}{\sum \left\| F_{PH_{ob}} \right| - \left| F_{PH_{calc}} \right\|}$ ob means observed, calc means calculated.

Figure of Merit = $|F(hkl)_{best}|/F(hkl)$

2.2.7 Model building and refinement

The final 2.3 Å electron density map was of good quality. Beside the main

chain was seen well connected, the side chains were quite obvious. Model building was carried out with the program O (Jones, 1991). The skeleton of the density map was created with the program GMAP in the PHASES package, which was displayed using O. The C_α atoms were positioned using the "baton" option in O, in which the baton was pivoted along the skeletonized main-chain density to locate the C_α positions. The polyalanine model was generated automatically based on the C_α trace model using the fragment-fitting option of O. With the location of seleniums, the amino acids Met 164 and Met 168 were first identified as they ought to be fairly close in the map. Based on this starting point, alanines were slowly changed to the correct residues of the sequence, which yielded the completed model. The 2.3 Å model was refined with CNS (Brunger, 1998). This model was used for rigid body refinement against the 1.8 Å native data. Iteratively, the $2F_o$-F_c and F_o-F_c maps were created to find the misfit of the model. Water molecules were picked from the F_o-F_c map at the 3.0 σ level and checked with the $2F_o$-F_c map at the 2.0 σ level. Positional and temperature refinement was repeatedly performed until the structure was refined to an R-factor of 0.19 and an R_{free} of 0.23 using reflections with $|F| > 3.0\ \sigma(|F|)$. The geometry of the final model was checked with PROCHECK (Laskowski, 1993) and all parameters were within acceptable ranges. The refinement statistics are listed in Table 2-3.

Table 2-3 Refinement parameters

Resolution range (Å)	8–1.8 (Native)
Reflections (working/test)	12,202/1,340
[1]R_{cryst} / R_{free}	0.23 / 0.27 (all reflections)
R_{cryst} / R_{free} [with $\|F\| > 3.0\ \sigma(\|F\|)$]	0.19/0.23
Asymmetric unit	One Hex1 molecule (M_r = 19.694 kDa)
Final model:	
Non-hydrogen atoms	1128
Waters	331
Average B-factors (Å^2):	
Protein	20.90
Waters	54.82
R.M.S.D. in bond lengths (Å)	0.005
R.M.S.D. in bond angles (°)	1.351

[1]*R*-factor=$\sum_{hkl}||Fo(hkl)| - |Fc(hkl)|| / \sum_{hkl}|Fo(hkl)|$.

2.3 Hex1 crystal lattice is required for woronin body assembly

2.3.1 Overall structure of Hex1

Hex1 has a two-domain structure consisting of two mutually perpendicular antiparallel β-barrels, Figure 2-6. The N-terminal barrel is made by six antiparallel strands, β1–β6, formed by residues 32–101. In addition, there is a 3_{10} helix (H1) formed by residues 37–40 between α1 and α2. The C-terminal domain of Hex1 forms another β-barrel composed of five strands (β7–β11) and two helices (H2 and H3). Both the helices are positioned between strands β9 and β10. Helix H2 is an α-helix formed by residues 137–145 and H3 is a 3_{10} helix consisting of residues 149–152. Both the N- and C-terminal domains have well-defined hydrophobic cores and the region between the two domains is also mainly hydrophobic.

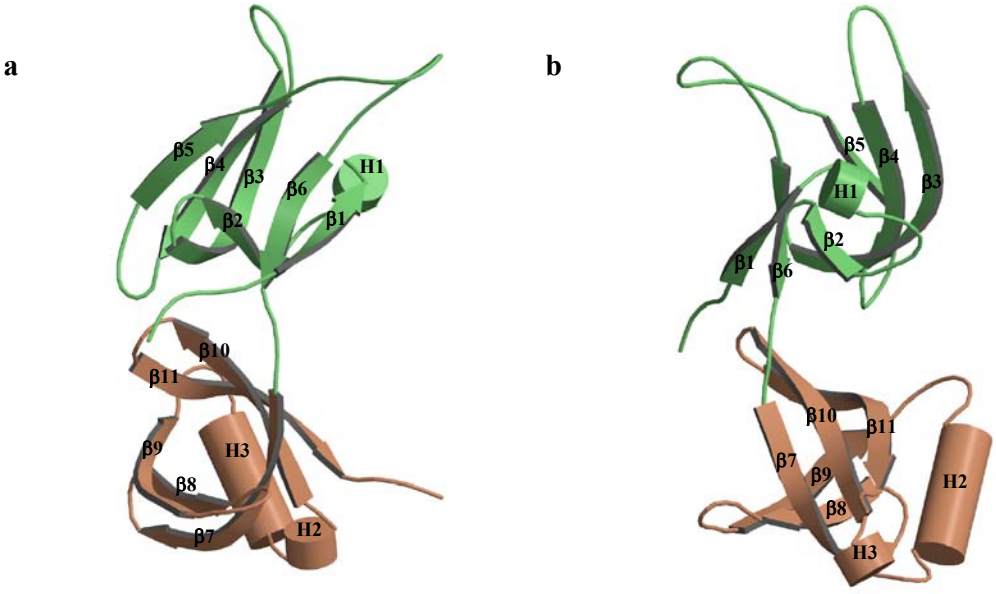


Figure 2-6. (a) Overall structure of Hex1. (b) Orthongonal view of (a).

Pictures are generated with Molscript (Esnouf, 1997) and Raster3D (Merritt, 1994).

2.3.2 Three groups of intermolecular interaction

During analysis of the Hex1 crystal lattice, three types of intermolecular surface interactions, designated Group I, II and III, were observed, Figure 2-7.

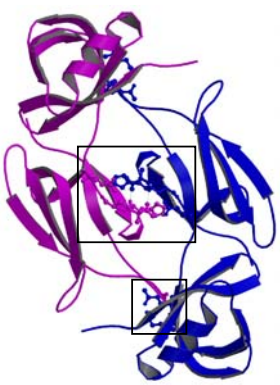
Molecules associated by the Group I interaction are related by a crystallographic two-fold rotation axis and their interaction is mediated by a central pair of salt bridges and a series of central and flanking hydrogen bonds, Figure 2-7a. Arg41 forms a salt bridge with Asp44, which is stabilized by a series of contacts with His39. The imidazole ring of His39 extends out just above the salt bridge, like a 'stabilizing hand', forming a π -interaction with the Asp44 side chain oxygen (Oδ2). Also, the carbonyl oxygen of His39 forms a hydrogen bond with the side chain nitrogen of Arg41 (NH1). The ends of the Group I dimer are linked by hydrogen bonds with the hydroxyl group and backbone nitrogen of Ser92, forming contacts with the backbone nitrogen and side chain oxygen (Oδ1) of Glu121, respectively. In addition, the carbonyl oxygen of Ser92 hydrogen bonds with the backbone nitrogen of Gln105, Figure 2-7b.

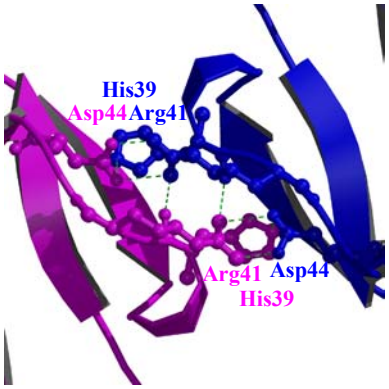
Amino acids making the Group II interaction (Figure 2-7c) are found on the opposite surface of Hex1 when compared with those determining the Group I interaction (compare Figures 2-7 a, c, 2-9). Molecules associated through the Group II interaction are related by the 6_5 screw axis, which generates the overall six-fold symmetry of the Hex1 crystal lattice. As in the Group I interaction, this interaction is also centered on a pair of salt bridges, which in this case is composed of Arg68 and

Glu81 (Figure 2-7c). These bonds are stabilized by two sets of hydrogen bonds: the carbonyl oxygen of Ser84 interacts with the side chain nitrogen of Gln49.

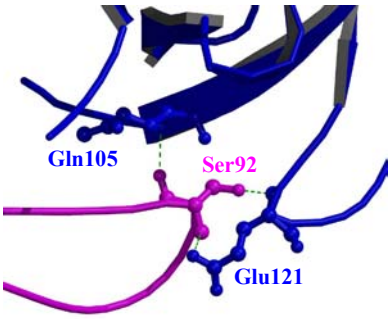
The Group III interaction is made up entirely of a series of hydrogen bonds that link two molecules through an asymmetric interaction (Figure 2-7d). The side chain nitrogen of Gln127 forms a hydrogen bond with the carbonyl oxygen of Ile56 and the backbone nitrogen and carbonyl oxygen of Val125 interact with the Gln134 side chain through a pair of hydrogen bonds. In addition to the three group contacts described above, several water-mediated hydrogen bonds are also observed in the Hex1 crystal lattice.

a

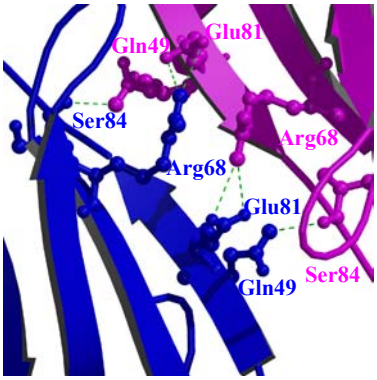



His39
Asp44
Arg41
Arg41
Asp44
His39

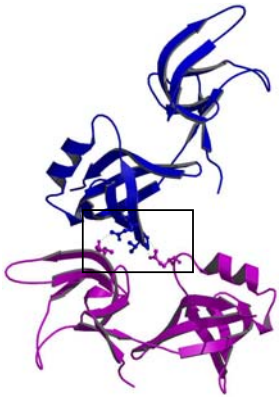
b


Gln105
Ser92
Glu121

c


Gln49
Glu81
Ser84
Arg68
Arg68
Glu81
Gln49
Ser84

d



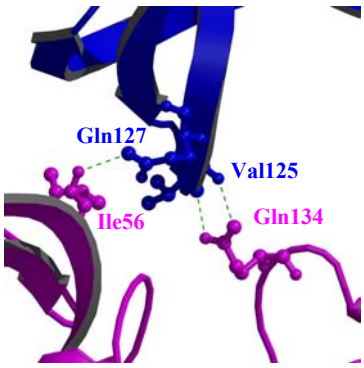

Gln127
Val125
Ile56
Gln134

Figure 2-7. Intermolecular interactions between Hex1 molecules. The left panels show the overall view of interacting molecules and their magnified views are shown in the right panels. In each case, one molecule is shown in blue and the other in purple. Interacting amino acids are labeled. (a) The Group I interaction is composed of the central N-terminal salt bridges (upper panel) and the N- to C-terminal domain contacts (lower panel). His39 plays a central role in stabilizing the salt bridge between Arg41 and Asp44. (b) The Group II interaction is localized to the N-terminal domain and centered on salt bridges between Arg68 and Glu81. Ser84 and Gln49 form a pair of flanking hydrogen bonds. The view of the dimer down the crystallographic *c*-axis reveals the 6_5 screw symmetry. (C) Group III interactions. The C-terminal amino acids Gln127 and Val125 of one molecule interact with N-terminal Ile56 and C-terminal Gln134 of the other molecule, respectively.

2.3.3 Interface of three groups of interaction

To prove whether the intermolecular interaction is of biological nature, one test is to check the interface of the contact. If the surface of two molecules is within 0.5 Å distance range, the surface is considered as an interface. Normally, if the interface is larger than 500 Å, the interaction is regarded as biological interaction (Hendrick and Thornton, 1998). Here in the Hex1 crystal, the Group I, II and III interfaces extend over 1,299, 697 and 515 $Å^2$ of surface area, calculated with Grasp (Nicholls, 1991), respectively, which are in the range of those associated with known oligomeric interfaces, Figure 2-8.

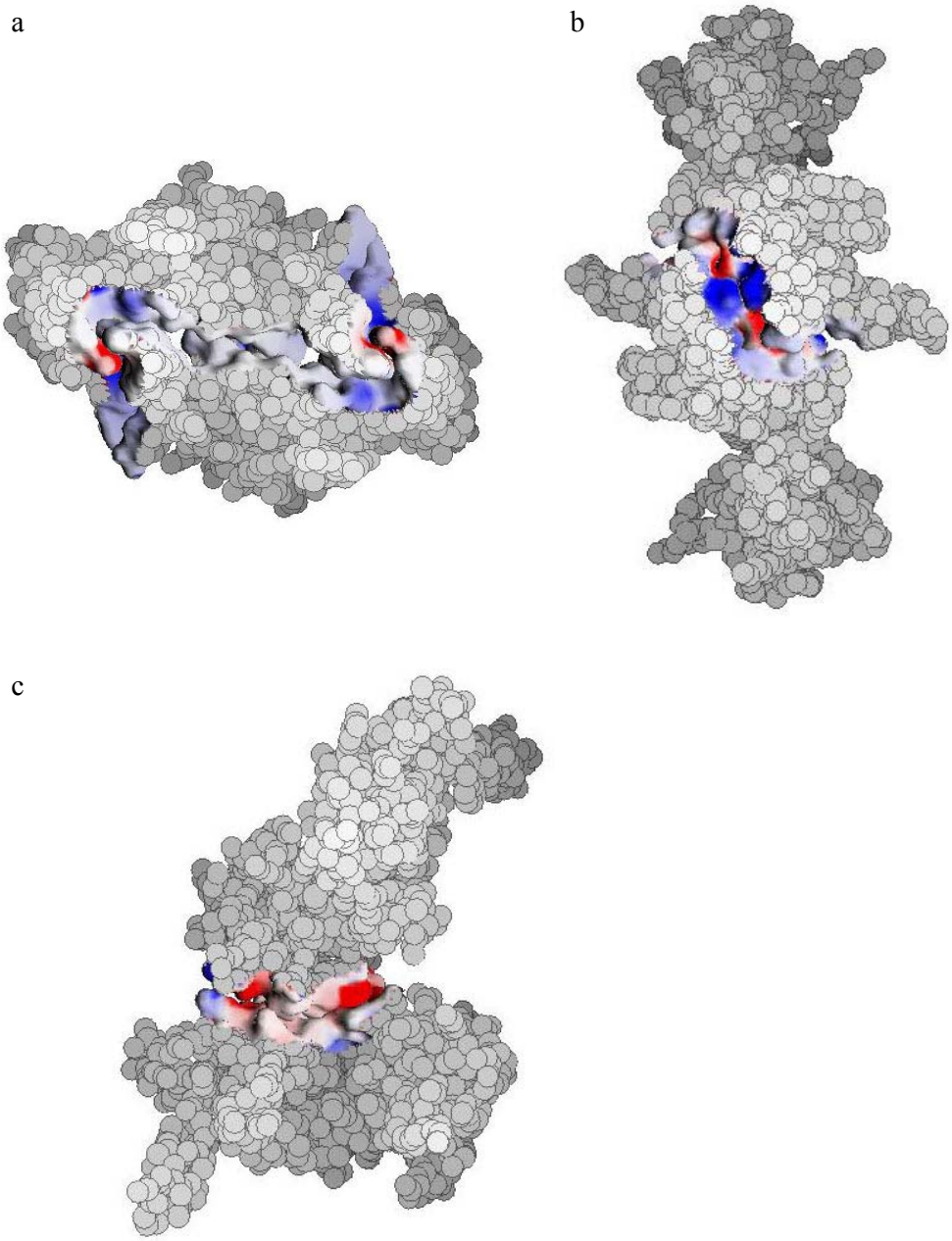
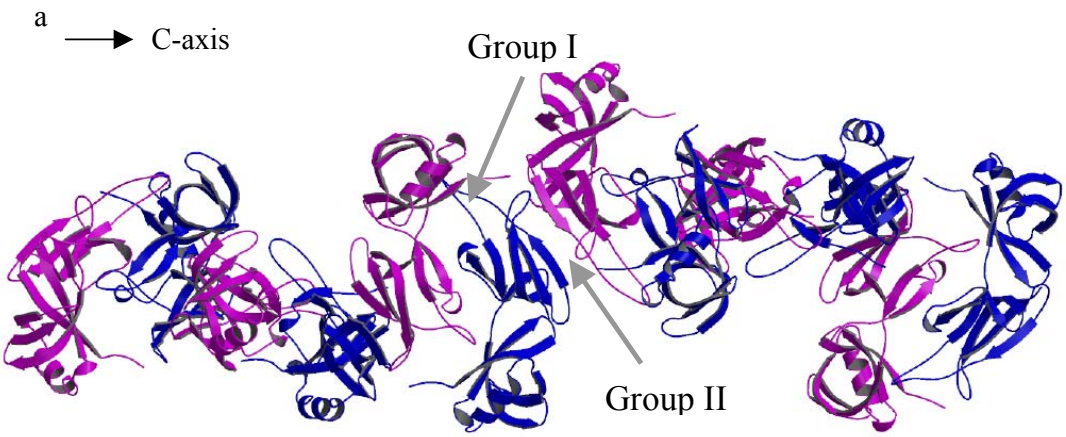


Figure 2-8. Interface of (a) type I (b) type II and (c) type III intermolecular interaction.

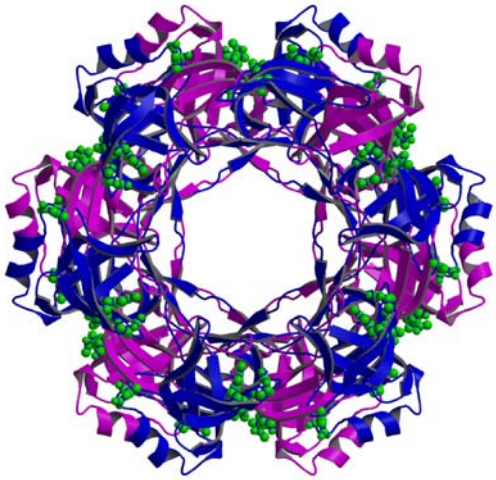
2.3.4 The packing of Hex1

Examination of the Hex1 crystal lattice reveals a structural polymer in the form of a helical spiral in which individual molecules are associated through alternating Group I and Group II interactions (Figure 2-9a). Rotation of the spiral is achieved by the crystallographic 6_5 screw symmetry, which rotates the associated molecules by 120° (Figure 2-7c, box). Thus, 12 Hex1 molecules (6 Group II mediated rotations) are required for each full turn of the spiral. Group III interaction residues are found on the surface of the Group I–Group II spiral (Figure 2-9b, Group III amino acids are colored green), where they are presented in 6_5 screw increments. This allows each filament to interact with six identical neighbours, producing the overall six-fold symmetry of the Hex1 crystal lattice, which is also seen in the native *N. crassa* Woronin body (Figure 2-9c, inset). Each spiral forms a tunnel with an inner diameter of 34 Å (Figure 2-9b). A portion of this tunnel is occupied by residues 1–26 of the Hex1 molecule, which are disordered and not visualized in the electron density map.

a



b



c

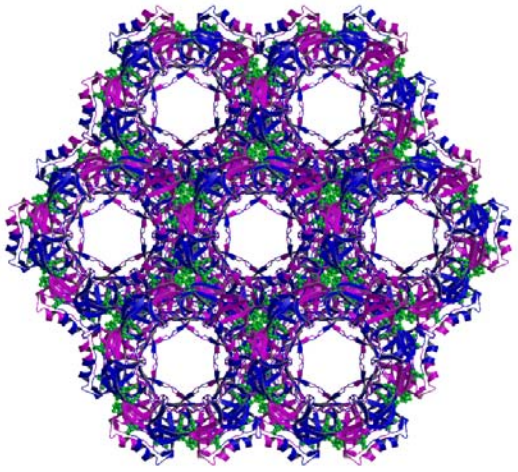


Figure 2-9. Molecular organization of the Hex1 crystal lattice. Alternating Hex1 molecules are shown in blue and purple. (a) Alternating Group I and Group II interactions (gray arrows, shown in the same orientation as in Figure 2-7) produce a coiled filament. The twisting of the filament is determined by the 6_5 screw axis (see Figure 2-6c, box) and 12 Hex1 molecules are required for one turn of the filament. The filament is viewed perpendicular to the c-axis. (b) The same coiled filament as in *(*a), viewed down the crystallographic c-axis. Group III interaction residues (green) are found on the filament surface. The tunnel in *(*b) is probably occupied by the first 26 N-terminal amino acids, which are missing from the electron density map. (c) Seven packed filaments reveal the overall Hex1 crystal lattice, which is shown next to a native *N. crassa* Woronin body (also viewed down the crystallographic c-axis, b*ar* = 1 μm).

2.3.5 Point mutations in Hex1 abort *in vitro* crystallization

To examine the role of crystal contact residues in Hex1 self-assembly, a series of mutants were prepared using the Quickchange site directed mutagenesis kit (Stratagene), Table 2-4. Their ability to self assemble was examined *in vitr*o. Mutations in the π-interaction residue His39 (H39G, Group I) and the hydrogen bonding residue Gln127 (Q127A, Group III) abolish the ability of Hex1 to crystallize *in vitr*o. In contrast, mutations in two residues that are not involved in intra- or intermolecular contacts, Lys143 (K143A, H2) and Arg149 (R149A, H3), do not interfere with self assembly (Figure 2-10a). To exclude the possibility that these defects in assembly were due to defects in Hex1 folding, these proteins were analyzed

by native PAGE. Both the assembly defective mutants, H39G and Q127A, migrated similarly to the wild type Hex1 protein in native PAGE (Figure 2-10b), indicating that these proteins are properly folded.

Table 2-4 Primer list

Constructs	Primer-sequence
Hex1-H39G	CATCCCCTGCCATGGCATCCGCCTCGGC
Hex1-K143A	GGAACCGTCTCCAGGCTGCTTTCGAGTCTG
Hex1-Q127A	GGCGATGTCAAGGCAAATTTGCCCGTCATTGAC
Hex1-R149A	CTTTCGAGTCTGGCGCTGGCTCCGTCCGTG

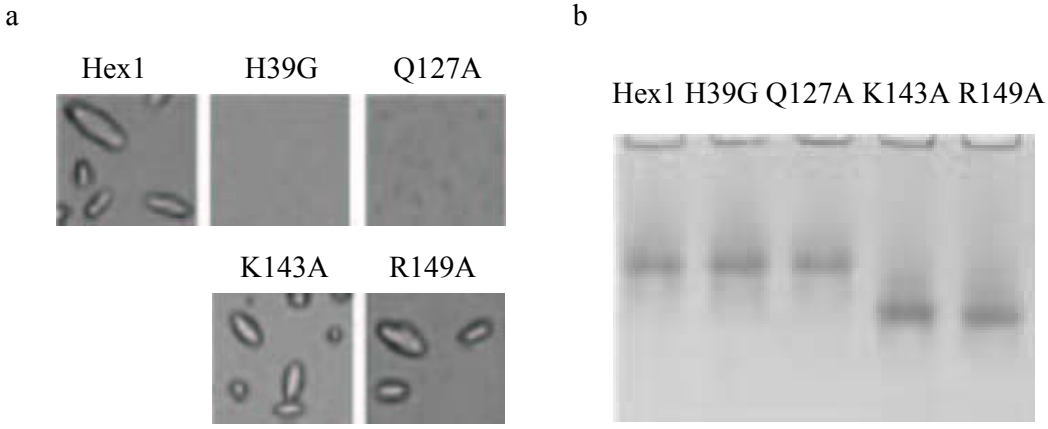


Figure 2-10. Requirement of crystal contact residues for Hex1 self assembly *in vitr*o. **(**a**)** The indicated recombinant Hex1 proteins were examined for their ability to self assemble *in vitr*o. These microcrystals are viewed perpendicular to the crystallographic c-axis. **(**b**)** The same proteins used for (a) were subjected to native PAGE. The increased mobility of the K143A and R149A mutant proteins is due to the change in their net charge.

2.3.6 Point mutations in Hex1 abort Woronin body formation

The mutation constructs H39G, K143A, Hex1 and empty vector were transformed to the *N. crassa* Hex1 deletion strain (ΔHex1:hph pan-2 pyr-4) by electroporation *(http://www.brinkmann.com/ELEC_appl_neurospora.asp)* and selected on minimal medium lacking pantothenate. The levels of Hex1 and mutant protein expression in these strains were examined by western blotting and determined to be equal for the strains shown. Expression of Hex1 or Hex1-K143A resulted in the production of the Woronin body with characteristic geometry, which could be observed at the cell periphery by light microscopy (Figure2-11a). In contrast, expression of Hex1-H39G (Group I mutant) resulted in the production of aberrant spherical Woronin bodies (Figure 2-11a). Electron microscopy further demonstrated that the Hex1-H39G Woronin body core is spherical and enclosed by a unit membrane (Figure 2-11d). Thus, the H39G mutation interferes with the intra-peroxisomal assembly of Hex1 but not with the PTS-1 mechanism of peroxisomal import. To further demonstrate the defective core assembly in the H39G mutant, cell extracts from Hex1 and Hex1-H39G expressing cells were treated with detergent to remove the Woronin body membrane and then subjected to differential centrifugation (Figure 2-11e). Wild type Hex1 readily sediments, suggesting it is a large and stable complex. In contrast, the Hex1-H39G mutant protein failed to sediment, confirming that it is either soluble or forms relatively small complexes. To assess the ability of these vesicles to substitute for Woronin body function, aerial hyphae, which extrude the protoplasm and produce fewer conidia (spores produced from aerial hyphae) in the absence of Woronin bodies (empty vector, Figure 2-11b, c) is examined. The expression of Hex1 or Hex1-K143A suppressed both the protoplasmic bleeding and conidiation defects of the deletion strain. However, the Hex1-H39G mutant was

unable to suppress either of these defects (Figure 2-11b, c). Thus, the H39G mutation interferes with normal Hex1 self-assembly (Figure2-11a, d, e) and abolishes Hex1 function (Figure 2-11b, c) even though Woronin bodies can grow to their normal size (Figure 2-11a, d).

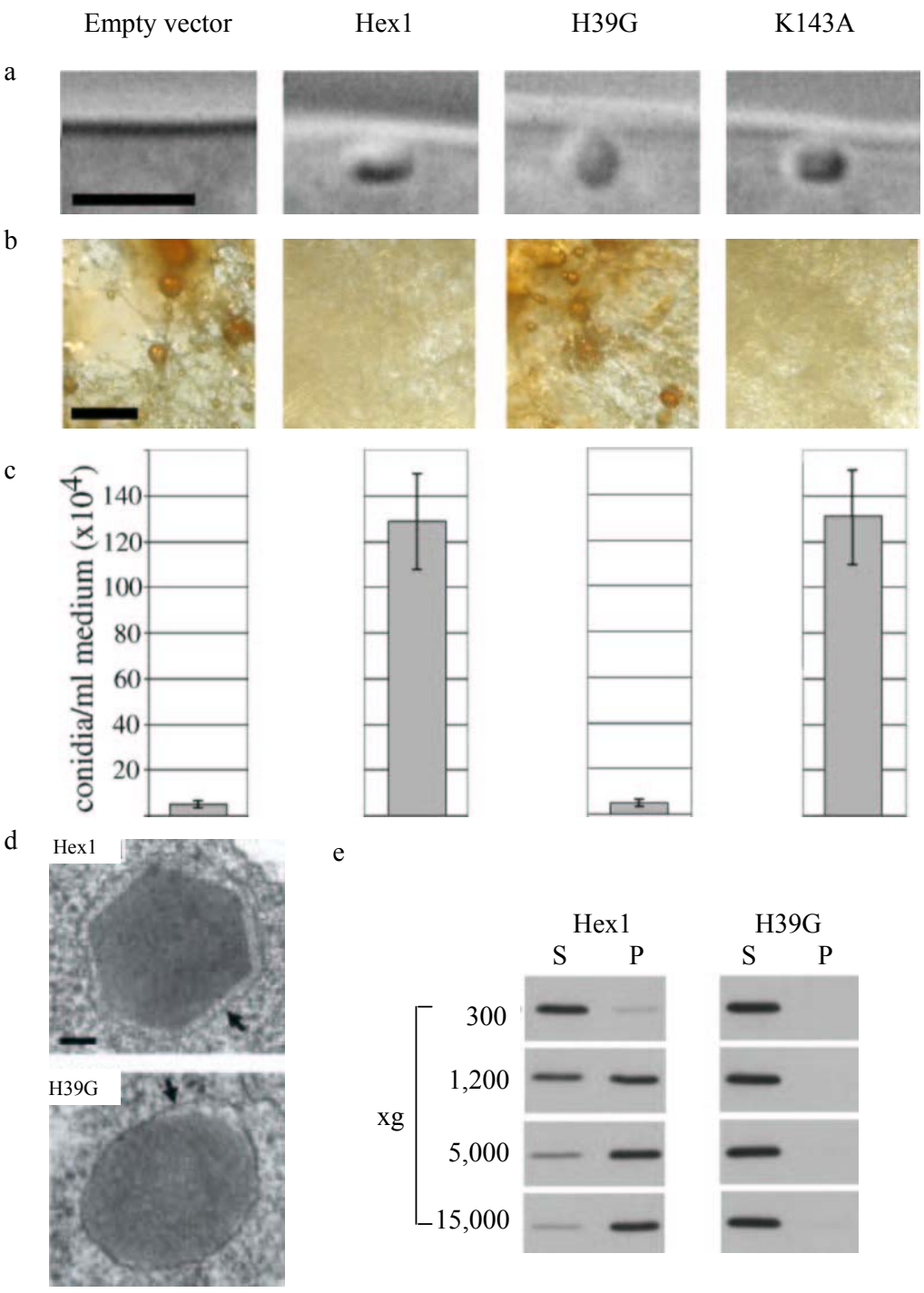


Figure 2-11. Requirement of crystal contact residues for Hex1 function. (a) Light microscopic examination of *N. crassa* transformants expressing the indicated proteins. Bar = 1 μm. (b) Microscopic examination of aerial hyphae and protoplasmic bleeding of cells grown

on solid medium. Bar = 1 mm. (c) Conidiation in the indicated strains grown on solid medium for 1 week. Values are the average of three independent measurements. (d) Representative Woronin bodies from the indicated strains visualized by electron microscopy. Arrows show the unit membrane of the vesicle. Bar = 100 nm. (e) Differential centrifugation of whole cell extracts from Hex1 and Hex1-H39G expressing cells. Extracts were centrifuged at the indicated centrifugal forces for 2 min and fractions were separated into pellet (P) and supernatant (S) fractions, which were analyzed for the distribution of Hex1 by western blotting.

2.4 EVOLUTIONARY ORIGIN OF HEX1

2.4.1 Hex1 structure homologs

Submission of the Hex1 coordinates to the Dali server (Holm and Sander, 1999) identified eIF-5A from *Pyrobaculum aerophilum* (PDB entry 1BKB, Peat, 1998) as a structural homolog of Hex1 *(*with a Z-score of 10). Comparison of the Hex1 structure with that of eIF-5A from *P. aerophilum* and *Methanococcus jannaschii* (Kim, 1998) reveals that these proteins share a similar structural organization (Figure 2-12). The N-terminal residues 27-102 of Hex1 aligns well with that of eIF-5A from *P. aerophilum* when compared using the SSAP server of CATH (Orengo, 1997; Pearl, 2000) with 19% sequence identity and 83% structural overlap with a score of 88.72. When fit with the corresponding region of eIF-5A, the N-terminal residues 33–85 of Hex1 superimpose with an r.m.s. deviation of 2.1 Å. A similar analysis shows that the C-terminal residues 103-173 of Hex1 are also homologous to the corresponding

region of eIF-5A. An SSAP search shows that 83% of these equivalent fragments could overlap and 15% of the sequence is identical, with a total score of 87.47 (Figure 2-12). In addition to the structural homology described above, Hex1 shows sequence similarity to the human eIF-5A protein, gi:9966867, blastp E-value = 3e –08, 26% identity, 52% similarity (Jedd and Chua, 2000).

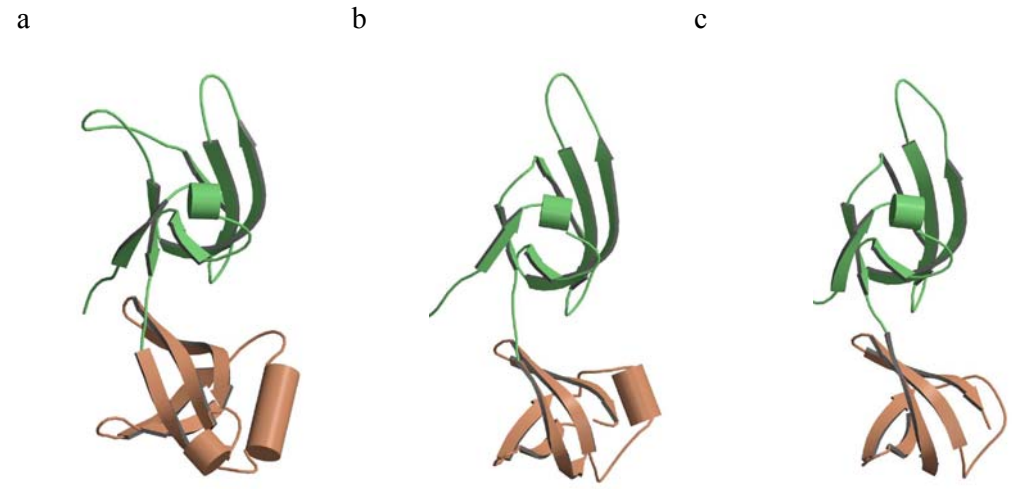


Figure 2-12. Overall structures of (a) Hex1 and eIF-5A protein from (b) *P. aerophilum* and (c) *M. jannaschii*. The three structures are depicted in the same orientation. All three proteins are composed of two mutually perpendicular antiparallel β-barrels. The N-terminal barrel is shown in green and the C-terminal region in brown.

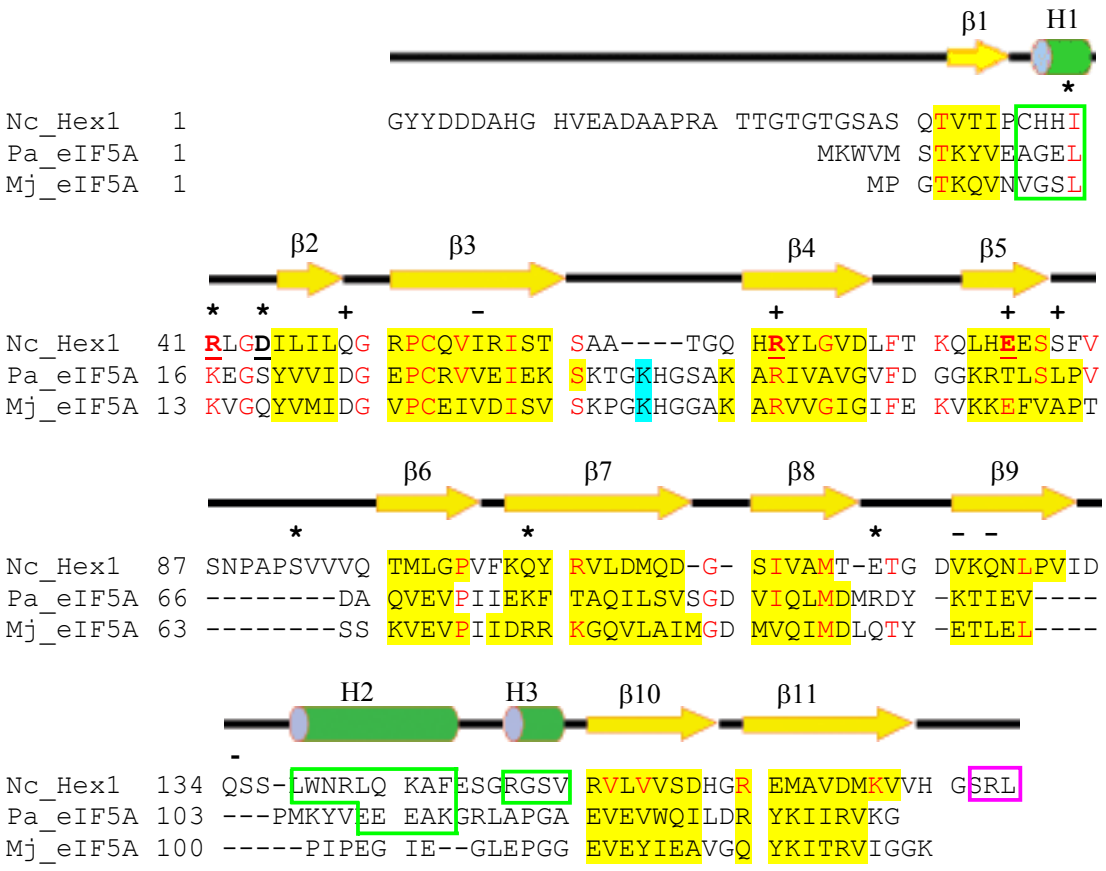


Figure 2-13. Alignment of the sequence of *N.* crassa Hex1 (Nc_Hex1) with that of *eIF-5A* from *P. aerophilum* (Pa_eIF5A) and *M. jannaschii* (Mj_eIF5A). The alignment is generated according to the SSAP output, as well as visual comparison. Secondary structure elements are boxed in green (helices) and highlighted in yellow (strands). Conserved residues are in red, and residues involved in the Hex1 Group I, Group II and Group III interactions are marked with asterisks, plus and minus symbols, respectively. The residues involved in forming salt bridges are underscored. The highly conserved lysine residue that is modified to hypusine in eIF-5A proteins is highlighted in cyan and the peroxisome targeting signal (PTS-1) of Hex1 is boxed in purple.

2.4.2 eIF-5A

eIF-5A is unique in that it is the only cellular protein so far known to contain the amino acid hypusine. Hypusine is formed through a spermidine-dependent posttranslational modification of eukaryotic initiation factor 5A (eIF-5A) at a specific lysine residue. The reaction is catalyzed by deoxyhypusine synthase (DHS) and deoxyhypusine hydroxylase (Xu, 2001). This modification is necessary for eIF-5A's activity.

eIF-5A was initially identified as a putative translation initiation factor based on its ability to stimulate methionyl puromycine synthesis under *in vitro* condition, but this has been questioned according to the fact that a similar effect on translation is not observed *in situ* (Wang, 2001).

Later work discloses that the interaction of eIF-5A with the Rev trans-activator protein of HIV mediates the transport of unspliced or incompletely spliced viral mRNA across the nuclear envelope (Schatz, 1998). In particular, the inhibition of eIF-5A blocked the nuclear export of nuclear export signals (NESs) derived from the HIV-1 Rev and human T cell leukemia virus type I Rex trans-activators, whereas the nucleocytoplasmic translocation of the protein kinase inhibitor-NES was unaffected (Elfgang, 1999). The immunofluorescence and immunogold electron microscopy study demonstrates that eIF-5A accumulates at nuclear pore-associated intranuclear filaments in mammalian cells and *Xenopus* oocytes. It interacts with the general nuclear export receptor CRM1. Furthermore, microinjection studies in somatic cells reveal that eIF-5A is transported from the nucleus to the cytoplasm which demonstrate that eIF-5A is a nucleocytoplasmic shuttle protein (Rosorius, 1999). *In vitro* binding studies demonstrate that eIF-5A is required for the efficient

interaction of Rev-NES with CRM1/exportin1 and eIF-5A interacts with the nucleoporins CAN/nup214, nup153, nup98, nup62 as well as nuclear actin (Hofmann, 2001).

eIF-5A has also been proposed to play a role in mRNA decay by acting downstream of decapping (Zuk, 1998). A recent report shows that DHS mRNA and eIF-5A mRNA show a parallel increase in abundance in senescing tissues and environmentally stressed tomato leaves and exhibiting programmed cell death, thereby suggesting the roles of eIF-5A in apoptosis (Wang, 2001).

2.4.3 Difference between Hex1 and EIF-5A

The function of Hex1 is clearly distinct from that of eIF-5A. First, Hex1 does not contain the highly conserved lysine that is modified to the amino acid hypusine, which is required for eIF-5A function (Schnier, 1991). Second, *Hex1* does not complement a yeast *eIF-5A* mutant (Jedd and Chua, 2000). Third, unlike the eIF-5A proteins, Hex1 is targeted to peroxisomes, where it self assembles into the dense core of the Woronin body (Jedd and Chua, 2000). The combination of sequence and structural homology between Hex1 and eIF-5A proteins strongly suggests that *Hex1* evolved *via the* duplication of an ancestral *eIF-5A* gene, followed by the acquisition of new functions. Purified mammalian eIF-5A exists as a dimer in solution and can reversibly associate to form tetramers and hexamers (Chung, 1991). However, eIF-5A crystal structures do not provide any evidence for such oligomerization (Kim, 1998; Peat, 1998). Thus, if eIF-5A self-assembles, it is probably through interactions distinct from those that mediate Hex1 assembly.

2.4.4 Selected Hex1 residues are highly conserved in eIF-5A

Comparison of the Hex1 group interaction residues with the analogous residues in eIF-5A proteins reveals that one of two residues associated with Group I and Group II salt bridges is invariably found in eIF-5A proteins (Figure 2-14). In addition, four out of seven intermolecular hydrogen bonds involve glutamine residues, and, in all cases, these correspond to conservative substitutions in the eIF-5A sequences, further suggesting an ancestral link between these two proteins. Finally, major structural differences between Hex1 and eIF-5A correspond to the extension of the loop between β5 and β6 and an extension of the C-terminal α-helix, H2. In both cases, these extensions contain residues, Ser92 in the case of the N-terminal loop (Figure 2-7a) and the nearby Gln134 in the case of α-helix H2 (Figure 2-7d) that participate in interactions. Thus, Hex1 self-assembly probably evolved through the exploitation of pre-existing residues, conservative amino acid substitution and changes in the length of secondary structural elements.

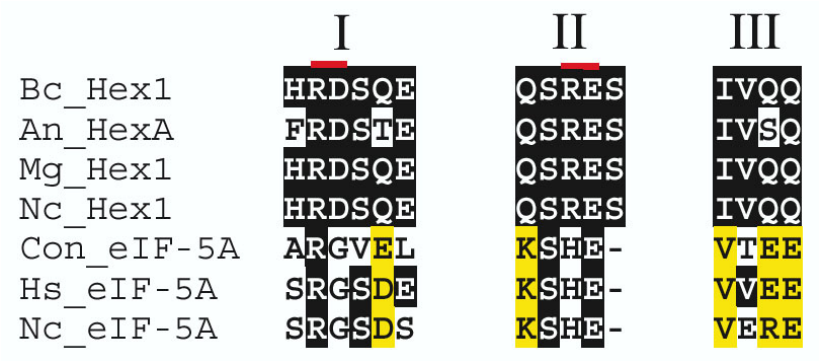


Figure 2-14. I, II, III alignment of *N. crassa* Hex1 (Nc_Hex1) crystal contact residues (Group I: His39, Arg41, Asp44, Ser92, Gln105, Glu121; Group II: Gln49, Ser61, Arg68, Glu81, Ser84; Group III: Ile56, Val125, Gln127 and Gln134) with Hex1 proteins from other Euascomycetes (Bc_Hex1, *Botrytis cinerea*; An_HexA, *Aspergillus nidulan*s; Mg_Hex1, *Magnaporthe grisea*) and the eIF-5A consensus sequences (Con_eIF-5A: N-terminal domain, pfam02869; C-terminal domain, pfam02187), as well as eIF-5A from *N. crassa* (Nc_eIF-5A; NCBI protein database entry gi:409567) and human (Hs_eIF-5A; gi:9966867). Residues shown are taken directly from a ClustalW alignment *(http://clustalw.genome.ad.jp/*) of the indicated proteins with *N. crassa* Hex1. Identical residues are shaded in black, and conserved residues are in yellow. The identity of the interaction Groups (I, II, and III) is shown above each block of sequence alignment. The red bar indicates salt bridge residues associated with the Group I and Group II interactions.

2.4.5 Evolutionary relationship between Hex1 and eIF-5A

Ascomycetes with filaments and complex reproductive structures were present 400 million years ago (Taylor, 1999), implying an ancient origin of Woronin bodies. Woronin bodies and *Hex1* have been reliably observed only in Euascomycetes (Markham and Collinge, 1987) where they are functionally (Jedd and Chua, 2000; Tenney et al., 2000) and physically (Markham and Collinge, 1987; Momany et al., 2002) associated with the septal pore. Euascomycetes form a clade within the ascomycetes. That is, they are all descendents of a single common ancestor and that ancestor did not give rise to any of the other ascomycetes (for example, *S. cerevisiae* and *S. pombe*, Berbee, 2001). Neither *S. pombe (http://www.sanger.ac.uk/Projects/S_pombe/*) nor *S. cerevisiae (http://genome-www.stanford.edu/Saccharomyces*/) genome appears to contain *Hex1* orthologs. This suggests that Woronin bodies arose in the ancestral euascomycetes. The restricted distribution of *Hex1,* its key role in a process characteristic of euascomycetes and evidence that it arose *via* gene duplication support the long held belief that such duplications and the consequent acquisition of new protein function may lead to key innovations underlying evolutionary success (Ohno, 1970).

2.5 DISCUSSION

Hex1 is the main component of the Woronin body and the crystal of Hex1 shares the same hexagonal prism profile as the native Woronin body, which leads to the question whether the Hex1 lattice shares the same lattice as the Woronin body? Several lines of evidence suggest that the Hex1 crystal lattice is a valid structural

model of the Woronin body core. Group I and Group II interactions are both centered around a bidentate arrangement of salt bridges that have been observed at oligomerization interfaces and are unlikely to occur in random crystal packing contacts. The Group I, II and III interfaces bury 1,299, 697 and 515 $Å^2$ of surface area, respectively, which are in the range of those associated with known oligomeric interfaces (Hendrick and Thornton, 1998). In addition, random crystal contact residues tend not to be conserved, whereas those associated with protein oligomerization are conserved (Elcock, 2001; Valdar, 2001). Sequence alignment of Hex1 orthologs reveals a high level of crystal contact conservation (Figure 2-14). Finally, a mutation that abolishes Hex1 crystallization also abolishes Hex1 function (Figure 2-11b, c) while not interfering with protein folding (Figure 2-11e) or vesicular targeting (Figure 2-11a, d). The mutants study shows that a Hex1 crystal lattice is absolutely required for Woronin body function in *N. crass*a.

The Hex1-H39G mutant, which abolishes the Hex1 crystal lattice, shows the Woronin body without a solid core. Why does the Woronin body require a solid core? Intercellular translocation of the protoplasm through septal pores is a common feature of fungal cells. Microscopic examination of living *Hex1* mutant hyphae shows that a single damaged hypha can extrude many cell equivalent of the protoplasm and large organelles such as vacuoles by distorting their shape as they pass through the septal pore. These observations suggest the presence of intra-cellular turgor pressure and in this case, only a dense-core vesicle is capable of sealing the septal pore.

Magnaporthe grisea is an ascomycetous fungus that causes the devastating blast disease in graminaceous hosts like rice, wheat, barley and millet. Similarily, a mutation in the *Magnaporthe grisea* Hex1 homolog has also resulted in disruption of self-assembly of woronin body as well as the morphological and functional defects in

appressoria (infectious structures), which finally delayed host penetration and subsequently disrupts invasive hyphal growth in *planta*. (Soundrararajan, 2004) According to the high homology and the conservation of all the contact residues between *Magnaporthe grisea* Hex1 and *Neurospora crassa* Hex1, it can be expected that *Magnaporthe grisea* woronin body share the samilar crystal lattice and interface. So chemical designed to disrupt the intermolecular interaction would has good commercial potential to combat the devastating blast disease in rice, wheat, barley and millet.

Although no evidence shows that eIF-5A can form a similar oligomer as Hex1, as seen from its crystal structure, the capacity that eIF-5A exists as a dimer and can reversibly form tetramers and hexamers may provide a rationale for an evolutionary selection of *eIF-5A* as a precursor to *Hex1*. EIF-5A self assembly might be enabled following its peroxisomal targeting, which is expected to significantly concentrate this normally cytoplasmic protein. Thus, peroxisome targeted eIF-5A may have assembled into a primitive Woronin body core. More work is necessary at the functional and structural level by expressing variant eIF-5A proteins in yeast and *Hex1* mutant fungi to test the above idea.

CHAPTER 3 DIMERIZATION OF GEMININ COILED COIL REGION IS NEEDED FOR ITS FUNCTION IN CELL CYCLE

3.1 INTRODUCTION

3.1.1 Discovery of Geminin

Geminin cDNA was discovered by looking for proteins that were degraded by mitotic, but not interface of *Xenopus laevis* egg extracts. In synchronized HeLa cells, this 25 kDa protein is absent during G1 phase, accumulates during S, G2 and M phases, and disappears at the time of the metaphase-anaphase transition (McGarry and Kirschner, 1998). A second Geminin cDNA was isolated from a screening procedure designed to identify proteins that affected *Xenopus* embryotic development. Its overexpression will expand the neural plate at the expense of the adjacent neural crest and epidermis (Kroll et al., 1998). As the proteins encoded by the above two cDNAs are 89% identical at the amino acid level, the protein was named Geminin after *Gemini*, which means “twins” in Latin. The protein has a calculated molecular mass of about 25 kDa, but they migrate aberrantly on polyacrylamides gels with apparent molecular mass of about 35 kDa (McGarry and Kirschner, 1998).

3.1.2 Role of Geminin in DNA replication

Every proliferating eukaryotic cell is faced with the prospect of having to copy

accurately and precisely its entire DNA in only a few hours during the cell cycle. Either incomplete replication or overreplication would cause cell death, or even worse, cause cancer in humans. To accomplish this process in the allotted time, eukaryotic cells have developed a 'divide and conquer' strategy. Unlike prokaryotic cells, eukaryotic genomes are replicated from multiple replication origins distributed along their chromosomes. Each replication origin is only responsible for the replication of a relatively small portion of the genome. This strategy allows the rapid replication of large genomes but brings with it a serious book-keeping problem. How can the cell keep track of all these origins, ensuring that each one fires efficiently during S phase while also ensuring that no origin has fired more than once? To cope with this, eukaryotic cells have evolved a remarkable molecular switch. Cyclin-dependent kinase and Geminin play key roles in the switch (Diffley, 2001).

Before eukaryotic DNA replicates, pre-replication complexes (pre-RCs) must be assembled at the replication origins in a reaction known as "licensing". Pre-RC assembles in a stepwise manner: origin recognition complex (ORC), a sequence-specific DNA binding protein complex, binds first and remains bound to origins during most or all of the cell cycle. The onset of S phase appears to be controlled by six proteins (MCM 2-7) that form MCM (minichromosome maintenance) complex. As cells exit from mitosis, the MCM complex interacts with chromatin and licenses DNA for replication. The fact that MCM family members have weak helicase activity shows that the complex is likely to open up chromatin and provide access for the enzymes that replicate DNA. The loading of MCM proteins onto chromatin is a key step in controlling the initiation of DNA replication. MCM loading requires initiating factor Cdc6/18 (Cdc6 is in *Saccharomyces. cerevisiae*, and Cdc18 is in *Schizosaccharomyces pombe*), which accumulates in the nucleus as cells exit mitosis and enter G_1. Another

initiating factor, Cdt1, is expressed when cells exit mitosis. Furthermore, it can form a complex with Cdc6/18. When Cdc6 or Cdt1 is depleted, MCM will not be able to load on the chromosome. Pre-RCs can only assemble at origins during a short period of the cell cycle between the end of mitosis and G1 phase. This temporal separation of pre-RC assembly and origin activation is a key feature of the switch because it ensures that new pre-RCs are unable to assemble on the origins that have already been fired. Thus, origins can fire just once in each cell cycle (Lygerou and Nurse, 2000).

There are at least two mechanisms identified on how DNA licensing is prevented after S phase. Cyclin dependent kinases (Cdks) are central to this regulation. Cdks are essential for triggering the initiation of DNA replication from origins that contain pre-RCs. At the same time, Cdks appear to play a direct role in preventing the assembly of new pre-RCs. In budding yeast, for example, Cdks target Cdc6 for SCF-dependent, ubiquitin-mediated degradation and trigger the export of the Mcm2-7 complex from the nucleus (Diffley, 2001). In mammalian cells, Cdc6 is phosphorylated by CDK2 and exported from the nucleus (Petersen, 1999; Saha, 1998). Because Cdk activity remains high from the onset of S phase until the end of the following mitosis, re-licensing of origins cannot occur until the beginning of the next cell cycle.

Another key regulator is Geminin. Working as a substrate of anaphase promoting complex/cyclosome (APC/C), Geminin is present in the cell nucleus from S phase until mitosis and is ubiquitinated and degraded as cells complete mitosis. The degradation requires a destruction box near its amino-terminus. The addition of Geminin to an *in vitro* replication assay containing *Xenopus* egg extracts blocks the association of MCM proteins with G1 chromatin, thereby inhibiting DNA replication.

The addition of Geminin blocks the initiation of DNA replication at the same stage as does the depletion of Cdt1 from *Xenopus* extracts. These findings were extended by the evidence that Geminin stably interacts with Cdt1 in *Xenopus* and human cells (Wohlschlegel et al., 2000; Tada et al., 2001). Geminin could therefore be important for repressing DNA licensing until the cell division is complete. Geminin is present in S phase cells where it may bind to and inactivate Cdt1 to ensure that DNA replication is not reinitiated at the origins that have just replicated. Geminin also accumulates in the nucleus of G2 cells, and could act redundantly to ensure the inhibition of S phase if licensing factors were to become inappropriately expressed in G2 cells. Geminin may also be important after DNA damage, halting S phase to give cells enough time to repair their DNA, figure 3-1 (Lygerou and Nurse, 2000).

Geminin does not only inhibit eukaryotic cell replication, but also inhibit the replication of cancer causing virus – Epstein-Barr virus (EBV) in eukaryotic cells. EBV, a 165 kb double stranded DNA virus of Herpes virus family, can replicate as an episome in latently infected cells. About 90% human B cells or epithelia cells are latently infected by EBV. This virus is highly related to carcinomas and lymphoproliferative diseases in immunosuppressed patients. Plasmids bearing the oriP replicator of EBV can not replicate in Δ/-Orc2 cell, where the Orc2 protein level is decreased about 90% compared to a normal cell. Reintroduction of Orc2 can rescue this defect. Interestingly, Geminin can efficiently inhibit the replication of this EBV plasmid. These data suggest that eukatyotic cellular replication initiation proteins are necessary to support EBV replication from oriP. Geminin based agents show some potential to be used to treat and prevent EBV-associated neoplasias (Dhar et al., 2001).

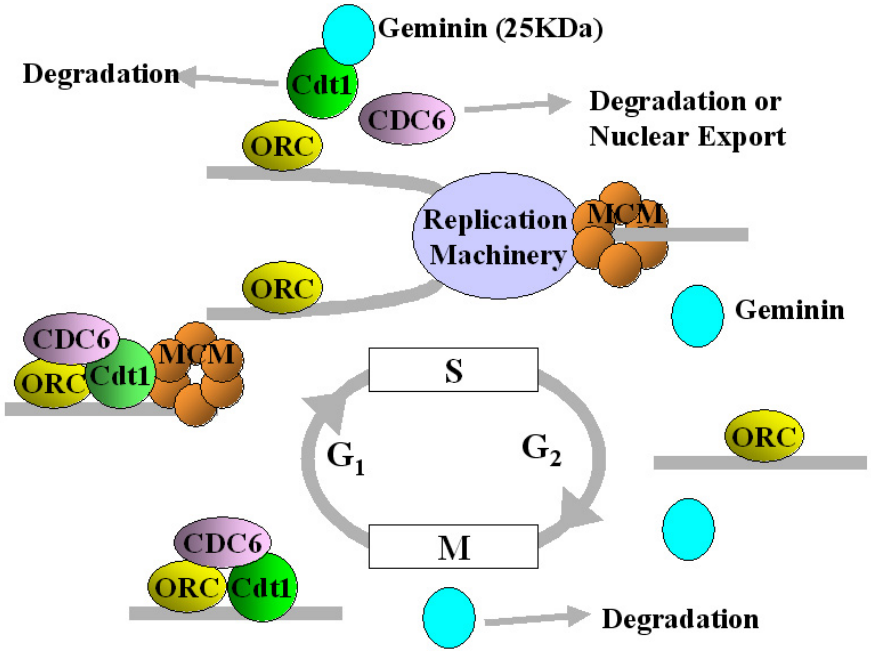


Figure 3-1. Role of Geminin in DNA replication

3.1.3 Role of Geminin in Neuron Differentiation

Besides the important role that Geminin plays in the cell cycle, it is also found to expand the neural plate at the expense of the adjacent neural crest and epidermis in *Xenopus* embryos when overexpressed. Maternal Geminin mRNA is found throughout the animal hemisphere from the oocyte through the late blastula. The misexpression of Geminin in the gastrula ectoderm suppresses BMP4 expression, the gene that promotes epidermial differentiation and blocks neural differentiation in ectoderm of both vertebrate and invertebrate, and converts prospective epidermis into neural tissue. Dominant negative Geminin lacking the neuralizing domain suppresses neural differentiation, and when misexpressed dorsally, produces islands of epidermal gene expression within the neurectodermal territory. This effect can be rescued by the coexpression of full length Geminin. Thus, Geminin seems to play an early role in establishing a neural domain during gastrulation in *Xenopus* embryo development

(Kroll et al., 1998).

Besides being present in *Xenopus*, a homolog of Geminin is also present in *Drosophila*. Like *Xenopus* Geminin, Dm Geminin also shows characteristics as a neural differentiation factor. The ectopic expression of Dm Geminin in the embryo can also induce some cells to ectopically differentiate as neural cells (Quinn et al., 2001).

3.1.4 Role of Geminin in apoptosis

Upon overexpression of Dm Geminin in *Drosophila* embryos, cells entering mitosis are arrested in metaphase and then undergo apoptosis (Quinn et al., 2001). When non-degradable Geminin (the destructive box of Geminin is mutated) is overexpressed in U2OS cell, cells show an early S phase arrest with high cyclin E and undetectable cyclin A levels. Consistent with high p53 and Cip1/Waf1 levels, profound loss of phosphorylated Rb and Ser15 of p53 being phosphorylated in Geminin overexpressed cells, it is indicated that the overexpression of Geminin has activated the ataxia telangiectasia mutated (ATM) protein and ATM and Rad3-related protein (ATR) checkpoint pathways. Down regulated cyclin A results in the induction of Cip1/Waf1 which in turn blocks further S phase progression and eventually triggers apoptosis (Shreeram et al., 2002).

3.1.5 Geminin depletion cause G2 phase arrest in *Xenopus* development

Embryos in which Geminin is knocked out by the antisense technique can complete early cleavage divisions normally but are arrested in G2 phase immediately after midblastula transition. This arrest requires Chk1, the effector kinase of the DNA replication/DNA damage checkpoint, as the overexpression of dominant negative

Chk1 can bypass the arrest. This result indicates that Geminin has an essential function and the loss of this function prevents entry into mitosis by a Chk1-dependent mechanism. Thus, Geminin may be required to maintain the structural integrity of the genome or it may directly down regulate Chk1 activity (McGarry, 2002).

3.1.6 Behaviour of endogenous Geminin

Geminin interacts with a DNA replication initiation factor, Cdt1, to suppress the initiation of DNA replication in a *Xenopus* egg extract based cell-free system, leading to the expectation that the protein acts as an inhibitor of cell proliferation. Immunohistochemistry and immunoblotting for Geminin, however, reveal that the protein is expressed specifically in proliferating lymphocytes and epithelial cells. Geminin is widely expressed in several malignancies and the number of Geminin-expressing cells is directly proportional to cell proliferation. Therefore, instead of being a suppressor of cell proliferation, Geminin expression is positively correlated with cell proliferation. Consistent with this observation, the transient overexpression of wild type Geminin in cancer cells in culture did not produce a cell cycle block. Only the expression of destruction box mutated Geminin, which is more stabilized, can arrest cells at the G1-S transition (Wohlschlegel et al., 2002).

Recombinant and endogenous forms of *Xenopus* Geminin behave differently from one another such that a significant proportion of endogenous Geminin escapes proteolysis upon exit from metaphase. During late mitosis and early G_1, the surviving population of endogenous Geminin does not associate with Cdt1 and does not inhibit licensing. Following nuclear assembly, Geminin is imported into nuclei and becomes reactivated to bind Cdt1. This reactivated Geminin provides the major nucleoplasmic

inhibitor of origin re-licensing during late interphase. Since the initiation of replication at licensed origins depends on the nuclear assembly, results suggest an elegant and novel mechanism for preventing the re-replication of DNA in a single cell cycle (Hodgson et al., 2002).

3.1.7 Domain organization of Geminin

Geminin shows pleiotropic effects when overexpressed. This 25 kDa protein plays key roles in the inhibition of DNA replication and induction of the uncommitted embryonic cell to differentiate neurons. How does this protein have two distinct functions?

Originally, previous work suggests that the structure of Geminin consist three parts, Figure 3-2. In human Geminin (HGeminin), amino acids 23 to 31 form the destruction box, 31 to 79 form the potential neuralization domain and 112 to 146 make the coiled coil domain with five heptad repeats that bind to Cdt1 and inhibits DNA replication (Thepaut et al., 2002).

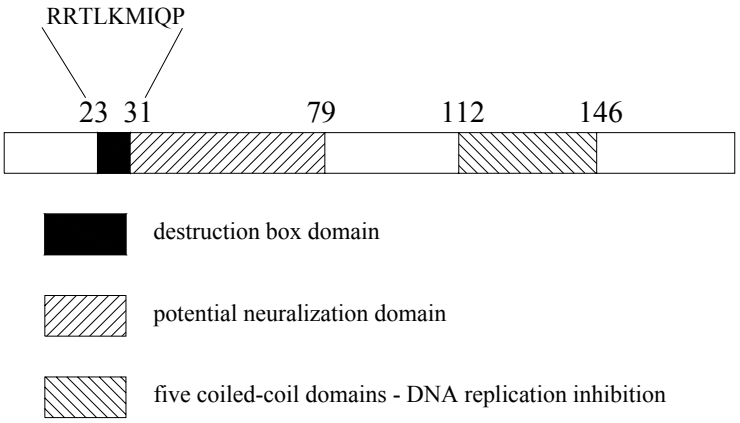


Figure 3-2. Domain organization of Human Geminin.

3.2 CRYSTAL STRUCTURE DETERMINATION OF GEMININ

3.2.1 Full length Geminin purification and Crystallization

Full length human Geminin was cloned into the pET-14b vector between the Nde1 and BamH1 sites with six histidines expressed at the N-terminus of the protein. The vector was transformed into BL21-DE3 cells. One colony was picked and used for the inoculation of 5 ml LB with 100 μg/ml Ampicillin for 16 h at 37 °C, 200 rpm. 5 ml of the overnight culture was added to 1 L LB with 100 μg/ml Ampicillin and continued to grow until OD_{600} was 0.6. Cells were induced to express His-Geminin with 1 mM IPTG. The cells were harvested after 3 hours by centrifugation at 6,000g for 10 min. The cell pellet was suspended in 10 ml lysis buffer D. The suspended slurry was passed through a French press for complete cell lysis. The lysate was then centrifuged at 20,000g for 30 min. to remove cell debris. The supernatant that contained His-Geminin was carefully decanted to a fresh 50 ml Falcon tube and then loaded to a 20 ml Econo-pac disposable chromatography column (Bio-Rad) with 3 ml pre-washed packed bead and mixed using a rotor wheel for 30 min. at 4 °C. The column was washed with 100 ml wash buffer E. The protein was eluted from the beads with 5 ml elution buffer F. As the protein eluted from Ni-NTA contained a small amount of DNA and traces of proteases might be wrapped inside, full length Geminin showed slow degradation. Hence, the fraction was further passed through a Hitrap-Q column (Amersham Pharmacia Biotech AB) to get rid of DNA, at 0.5 ml/min with FPLC (Bio-Rad). The column was first washed with 5 ml low salt solution, followed by 20 ml high salt solution, and again 5 ml low salt solution. Protein fractions were collected at 0.5 ml/tube with fraction collectors (Bio-Rad) and pure fractions were pooled together and concentrated to about 5 ml. The concentrated protein was further passed through a

Sephadex-200 column at 0.5 ml/min with FPLC to get rid of the traces of other contaminating proteins. The column was then washed with wash buffer G and 0.5 ml fractions were collected. The purity of the protein fraction was checked by running 15% SDS-PAGE gel (Figure 3-3a). The fractions which were pure enough for crystallization were pooled together and concentrated to 17 mg/ml and check with Bradford protein assay (Bradford, 1976). The protein was then aliquoted in 30 µl per tube, flash frozen with liquid nitrogen and stored at –80 ºC for later use. The protein was set up for crystallization with the Hampton Screen kit I and II by the hanging drop vapour diffusion method. Some tiny and bad quality crystals were obtained (Figure 3-3b). Further optimization of the crystallization condition was not successful.

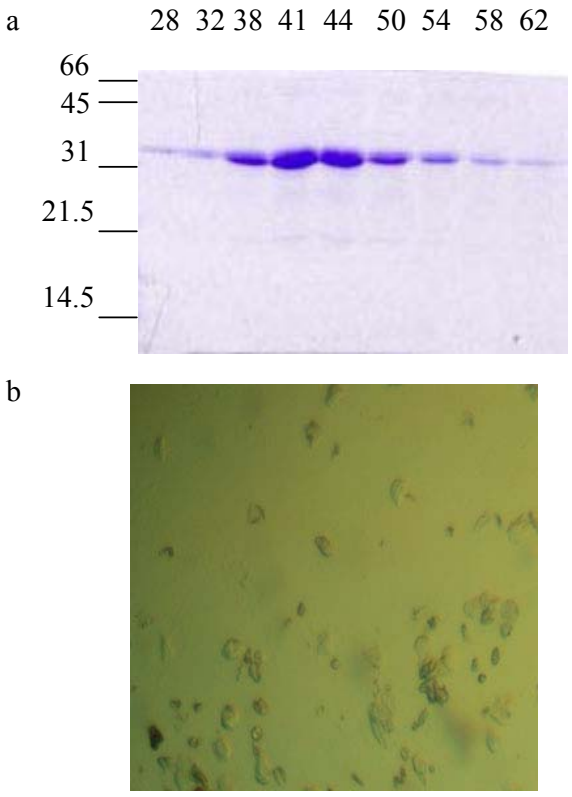
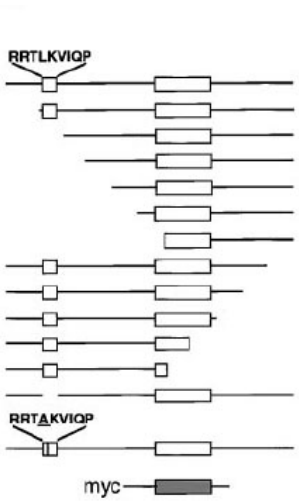


Figure 3-3. (a) Gel filtration fraction tube from sephadex-200. (b). Poor quality crystals of Geminin.

3.2.2 Identification of Cdt1 binding domain of Geminin

3.2.2.1 Past work on the domain study

Several truncated *Xenopus* Geminin constructs were made by McGarry (McGarry and Kirschner, 1998). According to the DNA replication test of those constructs, Geminin with the deletion of first 80 amino acids at the N-terminus can still inhibit DNA replication, while chopping off amino acids after residue 160 of the Geminin protein will not affect Geminin function (Figure 3-4).



MUTANT	HALF LIFE	PERCENT REPLICATION
WT	15 min	6
N30	15	8
N45	>90	2
N60	ND	4
N80	>90	5
N100	>90	119
N120	ND	113
C200	20	8
C180	10	12
C160	ND	18
C140	80	93
C120	15	89
DEL	>90	5
L36A	>90	ND
L 87-168	ND	3

Figure 3-4. Half life and DNA replication percentage of different Geminin mutants. The small square indicates the destruction sequence RRTLKVIQP, and the large rectangle indicates the predicted coiled coil region. The name of each mutant indicates whether the deletion is from the N or C terminus and the amino acid at which the deletion ends for an N mutant or where it starts for a C mutant. The half-life of each construct was measured with ^{35}[S]methionine labeled Geminin, incubated with the mitotic extract. Proteins were detected by electrophoresis and autoradiography. The percentage of control DNA replication occurring in the presence of each mutant (50–64 mg/ml) is also indicated. (McGarry and Kirschner, 1998).

3.2.2.2 Cdt1 binding study

Based on the fact that Geminin prevents DNA replication by interacting with Cdt1 and inhibits its function, proteins of the same constructs were translated by TNT Coupled Reticulocyte Lysate System (Promega) and then used to test Cdt1 binding ability (Figure 3-5). Besides N80 and C160 that bind to Cdt1, N100 can also bind to

Cdt1. According to McGarry's result, N80 can inhibit DNA replication while N100 can not, the result indicates that amino acids from 80-100 are critical for the inhibition of DNA replication, while the fragment 100-160 of *Xenopus* Geminin may be the minimum fragment required for binding Cdt1.

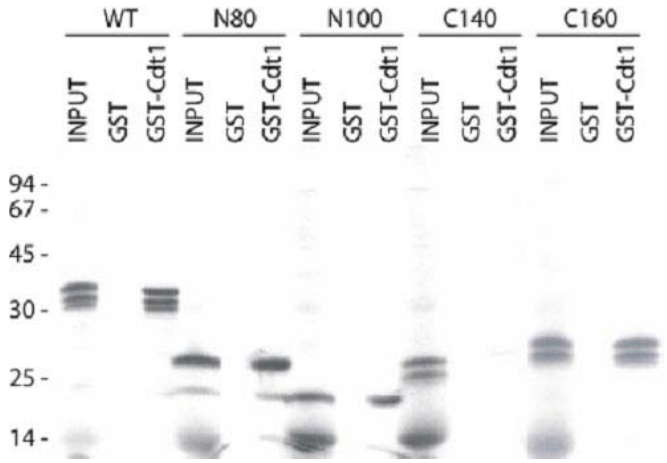


Figure 3-5. INPUT lane is the total IVT translation. GST lane is the IVT translated protein that binds to the GST bead. Cdt1 lane is the IVT translated protein that binds to the GST Cdt1 bead.

The sequence alignment of *Xenopus* Geminin with *Human* Geminin indicates that the equivalent segments of 80-160 and 100-160 of *Xenopus* Geminin are segments 70-152 and 93-152, respectively, in *Homo Sapiens* Geminin. The wheat germ extract IVT (Promega) translation proteins of these two segments were also used to test the Cdt1 binding ability as well (Figure 3-6). The result shows that both HGeminin70-152 and HGeminin 93-152 can bind to HCdt1.

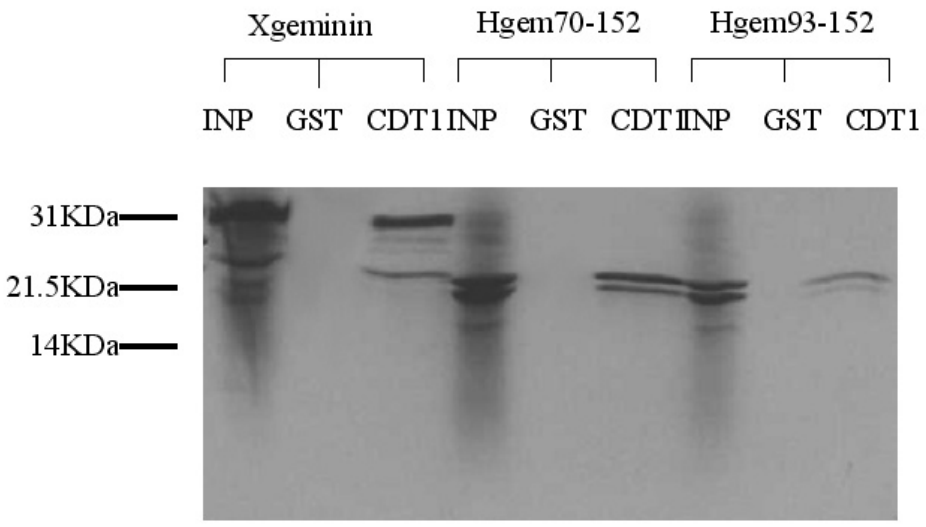


Figure 3-6. INP lane is the total IVT translation. GST lane is the IVT translated protein that binds to the GST-bead. Cdt1 lane is the IVT translated protein that binds to the GST-Cdt1 bead.

3.2.2.3 Function test of Geminin70-152

The transient replication of episomes based on *ori*P of Epstein Barr virus in mammalian cells is dependent on the components of Pre-RC and inhibited by the over-expression of Geminin (Chaudhuri, 2001; Dhar et al., 2001; Schepers, 2001). Furthermore, a bacterially produced plasmid is methylated on both strands of DNA and is susceptible to digestion by *Dpn*I. Since mammalian cells lack the *dcm* methylase, strands of DNA produced in mammalian cells are not methylated, so that the appearance of a *Dpn*I resistant plasmid in mammalian cultures is indicative of the replication of the bacterial plasmid. The ability of replication inhibition of HGeminin70-152 and HGeminin 93-152 is carried out by our collaborator with the above system. Constructs pEBG-HGeminin70-152 and pEBG-HGeminin93-152 were co-transfected with the Epstein Barr virus (EBV) plasmid p367 to the HCT116 cell line.

Assay of the transient replication of plasmid p367 was done as described (Dhar et al., 2001). The result shows that pEBGHGeminin70-152 successfully inhibits EBV DNA replication while pEBGHGeminin93-152 can not (Figure 3-7).

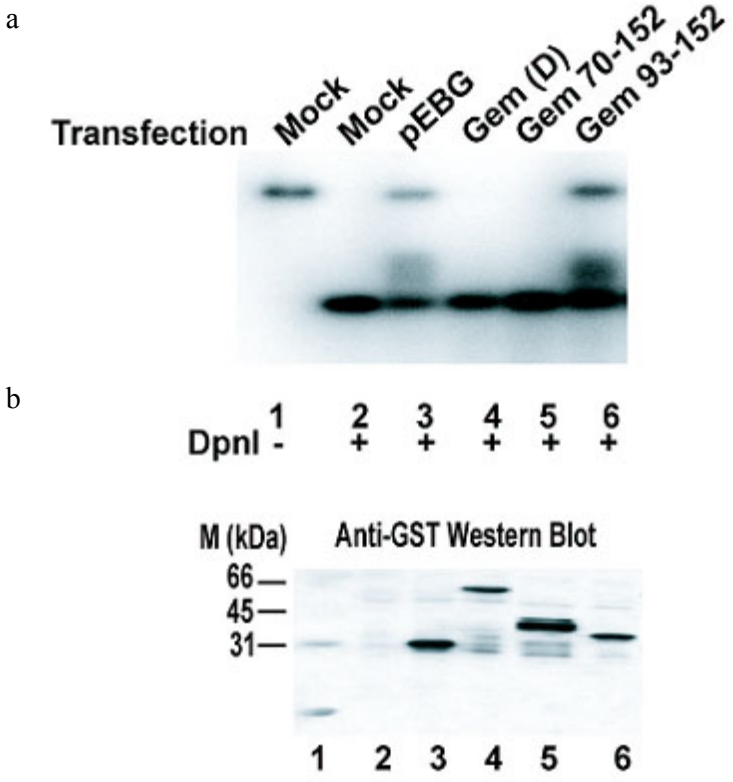


Figure 3-7. (a) HGeminin70-152 can inhibit Epstein Barr Virus oriP mediated replication whereas HGeminin 93-152 can not. Plasmids for GST (lane 3), GST-Geminin with the destruction box deleted [Gem (D), lane 4], GST-Geminin70-152 (lane 5) and GST-Geminin 93-152 (lane 6) were co-transfected with plasmid p367 in the HCT116 cell line and DpnI resistant p367 DNA was detected by southern blot. (b) Western blot analysis to show the expression of different constructs used in (a). Transfected cell lysates are after SDS-PAGE and anti-GST western blot.

Loading order in lanes 1-6 is the same as in (a).

3.2.3 Expression and purification of Geminin70-152

3.2.3.1 Expression and purification of Geminin70-152

The human Geminin70-152 fragment was amplified by PCR from full length Geminin and cloned into the pET28a vector between the BamH1 and Not1 sites. 5 ml LB with 25 μg/ml Kanamycin was inoculated with BL21::pET28a-Geminin70-152 and grown overnight at 200 rpm 37 °C. In the next morning all 5 ml culture was added to 1 Liter LB with 25 μg/ml Kanamycin and allowed to grow at 200 rpm 37 °C till OD 600 was 0.4 ~0.6. The protein was induced with 1 mM IPTG. Three hours later the cells were harvested by spinning at 6,000 g for 10 min. The cell pellet was resuspended in 10 ml lysis buffer H. The suspended slurry was passed through a French Press for complete lysis. The lysate was cleared by spinning at 20,000 g for 30 minutes. 3 ml of the Ni-NTA beads were pre-washed with 60 ml lysis buffer H to get rid of ethanol and the supernatant from bacterial lysates bound at 4 °C for 30 min. The beads were washed with 100 ml wash buffer I. Finally, the protein was eluted with 5 ml elution buffer J. 1 mM DTT and 1 mM EDTA were added to the eluted protein, which was pure enough for setting up for crystallization. The protein was concentrated to 14 mg/ml, frozen with liquid nitrogen and stored at –80 °C freezer for later use, Figure 3-8.

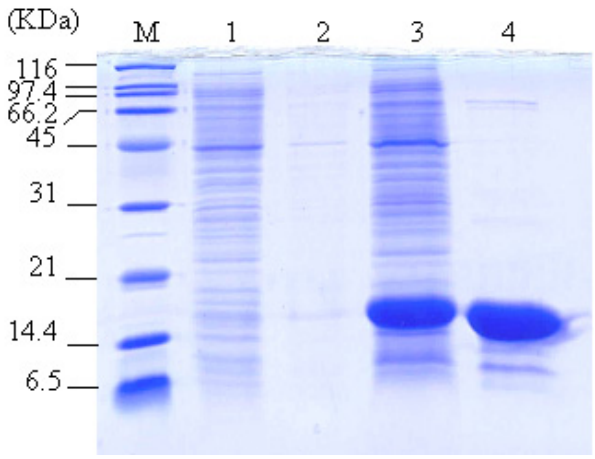


Figure 3-8. Native HGeminin70-152 purification. M is the marker. Lane 1 is the portion of the lysate that does not bind to the Ni-NTA beads. Lane 2 is the flow through after washing the Ni-NTA beads. Lane 3 is the insoluble portion of the total lysate. Lane 4 is the elution fraction after purification.

3.2.3.2 Expression and purification of Geminin70-152 containing selenium

One colony of fresh Geminin70-152 containing bacteria was inoculated to 5 ml LB with 25 μg/ml Kanamycin and grown overnight at 37 °C, 200 rpm. In the next morning the cells were gently spinned down at 1,300 g for 3 minutes, the pellet was washed twice with minimum medium M9 and added to 50 ml of the same, pre-warmed medium. The bacteria were allowed to grow at 37 °C 200 rpm until OD_{600} was 0.6 (about 8 h). The 50 ml culture was added to 1 L pre-warmed M9 medium and continued to grow until the cells reached OD_{600} of 0.6 to 0.9, which normally took about 24 hours. At this point, amino acids: lysine, phenylalanine, and threonine at 100 mg/l, isoleucine, leucine and valine at 50 mg/l and L- selenomethionine at 60mg/l were added to the culture. Fifteen minutes later, 400 μl of 500 mM IPTG was added to the culture to induce Geminin expression and the culture was harvested 6 h later.

The method of selenomethionine protein purification was the same as that of native Geminin70-152. The mass of the purified protein was checked with Liquid Chromatography Mass Spectrometer API300 (MDS SCIEX). The mass of Geminin70-152 and selenium substituted Geminin70-152 are 13,163 Da and 13,394 Da, respectively. The mass difference suggests that all five methionines in Geminin70-152 were replaced by selenomethionine (Coligan et al., 1995). The pure Selenium Geminin70-152 protein was concentrated to 14 mg/ml and flash frozen with liquid nitrogen before storage at –80 °C, Figure 3-9.

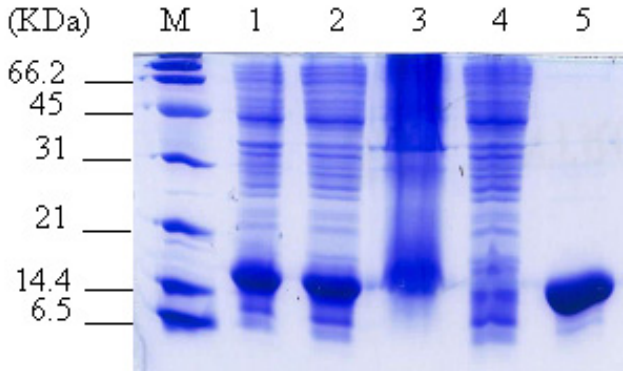


Figure 3-9. Selenomethionine HGeminin70-152 purification. M is the marker. Lane 1 is the total cell lysate. Lane 2 is the soluble fraction. Lane 3 is the insoluble fraction. Lane 4 is the portion that can not bind to the Ni-NTA bead. Lane 5 is the final elution of HGeminin70-152 selenomethionine protein.

3.2.4 Crystallization of native and selenomethionine Geminin70-152

Both native and selenium HGeminin70-152 proteins were set up for crystallization with the Hampton Screen kit by the hanging drop vapour diffusion method. 1 µl of the protein was mixed with equal amount of reservoir buffer. Both native HGeminin70-152 and selenium HGeminin70-152 crystals appear at the same condition consisting of 50 mM Tris HEPES (pH 8.5), 500 mM NaCl, 5% glycerol. Three days later, native HGeminin70-152 crystal could grow to 0.15 x 0.10 x 0.10 mm (Figure 3-10a) and selenium HGeminin crystal grew to 0.50 x 0.05 x 0.05 mm (Figure 3-9b).

a

b

Figure 3-10. (a) HGeminin70-152 and (b) Selenium HGeminin70-152 crystal.

3.2.5 Crystal data collection and processing

All crystals were cut to a proper size about 0.3 X 0.05 X 0.05 mm with micro-tools (Hampton Research). The crystals were flash frozen with liquid nitrogen in a cryoprotectant containing 50% (v/v) mineral oil and 50% paratone. Data for both native and selenomethionine crystals were collected at the beamline X12B, National synchrotron Light Source, USA. The native HGeminin70-152 crystal could diffract X-rays to a resolution of 2.0 Å, while the selenium HGeminin70-152 crystal diffracted X-ray to 3.0 Å. All diffraction frames were processed with DENZO 1.96 and scaled using SCALEPACK (Otwinowski, 1997).

Both crystals belong to the orthorhombic space group $P2_12_12_1$ and have unit-cell parameters a = 37.16 Å, b = 94.46 Å, c = 102.56 Å, Table 3-1. There are two molecules in the asymmetric unit. The solvent content is 65.3%.

Table 3-1 Native and Selenomethionine crystal data

Spacegroup:	$P2_12_12_1$			
Unit-cell:	a = 37.16 Å, b = 94.46 Å, c = 102.56 Å			
	$\alpha = \beta = \gamma = 90°$			
	Se-edge	Se-peak	Se-remote	Native
Wavelength (Å)	0.9789	0.9783	0.9611	0.9782
Resolution (Å)	2.8	2.8	2.8	2.0
No. of reflection	667,899	663,322	770,237	536,955
Unique reflections	9,950	10,061	10,535	25,309
Completeness (%)	99.7	99.7	99.2	99.7
R_{sym}	0.102	0.102	0.118	0.085

3.2.6 Model building and refinement

Two of the five selenium positions in each protein chain were determined by the MAD method with the SOLVE program, using 3.5 Å data (Terwilliger, 1999). The figure of merit of the solution was 0.69. Phases were calculated with the PHASIT module of the PHASES package, using 4.0 Å MAD data (Furey, 1997). To reduced featureless densities of the solvent, the phases were improved with 50 cycles of solvent flattening using the doall.sh script in the PHASES package. The electron density map calculated from the solvent flattened phases showed substantial improvement. The map was clear enough to show two long coiled coils. Phases were slowly extended to reflections upto resolution 3.0 Å with the extnd.sh script of the PHASES package. Electron density maps at various resolution limits were calculated and displayed in O (Jones, 1991) for model building. Side chains were well resolved in the 3.0 Å map and the selenium positions were recognized as Met151 of the parallel coiled coils. The structure was manually built by fitting in amino acids from the C-terminus to the N-terminus. The electron density before amino acid 92 was not clear. The determined structure was then fitted onto 2.0 Å native data by doing rigid body refinement with the CNS package (Brunger, 1998). The model was slowly refined by adjusting the improper regions according to the $2F_o$-F_c and F_o-Fc maps and performing positional and temperature refinement, until the final R-factor reached 0.216 (R_{free} = 0.239). The geometry of the molecule was judged to be well within acceptable limits when checked by PROCHECK (Laskowski, 1993). The refinement parameters are tabulated in Table 3-2.

Table 3-2 Refinement parameters

Phasing power:	
Dispersive Acentrics	3.67
Anomalous acentrics	5.93
Mean figure of merit (F.O.M)	0.762
Refinement parameters:	
Resolution range (Å)	8–2.0 (Native)
Reflections (working/test)	18,192/1,966
R_{cryst} / R_{free}	0.216 / 0.239 (All reflections)
Asymmetric unit	Two his-Geminin70-152 molecule (M_r = 13.2 kDa)
Final model:	
Non-hydrogen atoms	1034
Waters	120
Average B-factors (Å^2):	
Protein	35.6

Waters	73.45
R.M.S.D. in bond lengths (Å)	0.005
R.M.S.D. in bond angles (°)	0.90

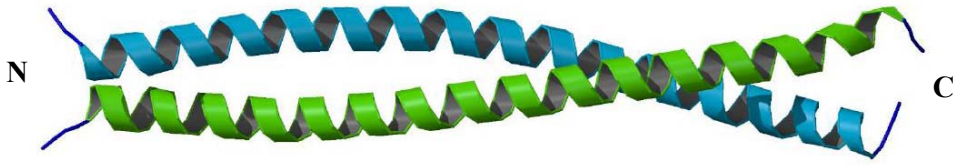
3.3 STRUCTURE OF GEMININ COILED COIL DOMAIN

3.3.1 The overall structure of Geminin coiled coil domain

The experimentally phased electron density clearly revealed residue 92-152. Amino acid 94 to amino acid 150 forms a continuous parallel coiled coil with another molecule. A pseudo 2-fold axis relates the subunits of the dimer. The overall structure of the peptide is a typical coiled coil with two extended α-helices that pack together in a left handed superhelix, forming a twisted elliptical cylinder, ~ 90 Å long and ~ 25 Å wide (Figure 3-11 a). The average distance between the helical axes is 10.4 Å.

N-terminal residues 70-91 were not visible in the electron density map, calculated at different ranges of resolution. To determine whether these residues were lost by proteolysis, SDS-PAGE (Figure 3-11 b) and Mass Spectroscopy were used to analyze the Geminin70-152 peptide in the crystals. The mass of the peptide in the Geminin70-152 crystal was 13162.64 Da and that in the Selenium Geminin70-152 crystal was 13394.44 Da, suggesting that the crystallized peptides were intact without any degradation. Therefore, the region immediately N-terminal to the coiled coil domain of Geminin is either unstructured or forms an ensemble of structures that are statically disordered in the crystal. This region is necessary for the inhibition of DNA replication and may only be stabilized by interacting with Cdt1.

a



b

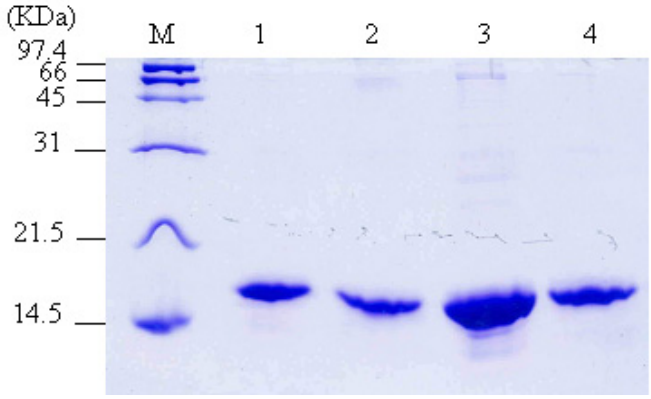


Figure 3-11. (a) The structure of Geminin92-152. (b) No degradation of HGeminin70-152 during crystallization. M is the marker. Lane 1 is the purified HGeminin70-152 protein. Lane 2 is the HGeminin70-152 crystal. Lane 3 is the Selenium HGeminin70-152 protein. Lane 4 is the Selenium HGeminin70-152 crystal.

3.3.2 Inter-subunit interactions of the Geminin coiled coil domain

Although the Cdt1 interacting domain of Geminin has the structure of a typical coiled coil, the sequence of the interacting surface is rather unusual. Coiled coil proteins are characterized by a repeating pattern of seven residues, $(abcdefg)_n$, with hydrophobic amino acids predominating at positions a and d of the heptad repeat (Burkhard, 2001). The a and d positions define a stripe of residues along the side of an α-helix that is buried by packing the two α-helical subunits together. In the coiled coil, the side chains of residues at a and d (the knobs) are inserted into spaces (the holes) between the side chains of the apposing α-helix to create a "knobs into holes" packing of residues at the dimer interface. That is the side-chains are idealized into uniform knobs which are then made to pack into the holes (the spaces between knobs) of the adjacent helix. (Crick, 1953).

The oligomeric state of a bundle of α-helices is strongly influenced by the shapes of the nonpolar side chains at the a and d positions of the constituent helices. The d positions are normally depleted of β-branched residues and leucine is often prominently featured at this position. In contrast, there is an abundance of β-branched residues like isoleucine or valine at the a position (Harbury et al., 1993, 1994). The side chains of the residue pairs at the a position extend toward the opposite sides of the dimer and pack with their C_α-C_β bonds aligned in a parallel orientation. The β-methyl groups of isoleucine or valine side chains are readily accommodated in the dimer interface by this parallel packing. The perpendicular packing at position d instead points the C_α-C_β bonds of interacting residues toward the dimer interface, resulting in less space for the β-methyl group of isoleucine or valine that usually occupies this position (Chang, 2003).

Although the Geminin coiled coil exhibits the usual knobs into holes packing at the dimer interface, its amino acid sequence does not conform to the consensus for coiled coil dimers (Figure 3-12b). Instead of β-branched amino acids occupying the a position, five out of seven amino acids are unbranched amino acid. There are two asparagines, one alanine, one leucine and one serine. Although nonpolar amino acids usually occupy the d position, two positively charged amino acid, Arg106 and Lys127 are present at this position in the Geminin coiled coil. The polar atoms of the side chains from Arg106 and Lys127 point away from the dimer interface and towards solvent with the carbons of the side chain providing the interaction surface. Previous work suggests that buried polar residues in coiled coils can be the important determinants of structural uniqueness, influencing both oligmeric states and strand orientation (Akey et al., 2001), and contributing considerably to the stability of coiled coils (Burkhard, 2001). Beside an alanine, a bulky amino acid, Trp99, is present in the d position (Figure 3-12c). Such a bulky residue is very rare in the interface of homodimers, because large residues (such as Phe, Tyr and Trp) could produce packing constraints in the interface (Figure 3-12a, b; O'shea, 1991). The existence of positively charged and bulky amino acids may destabilize the interface and suggests the possibility that the dimerization state of Geminin may be reversible *in vivo* in different physiological conditions.

Consistent with the previous data that charged residues are frequently observed at the e and g positions of coiled coils, residues occupying the e and g positions in Geminin are mainly charged. Four of seven residues at the g position are glutamic acid. In the e position, there are five positively charged residues and one negatively charged residue. These charged residues contribute to the stability of the coiled coil structure by forming broad ranges of intermolecular and intramolecular interaction.

His121 Nε2, Asp128 Oδ1, Lys135 Nζ of one chain and Glu116 Oε2, Lys127 Nζ, Glu130 Oε2 of the other chain form three pairs of salt bridges, respectively. These interactions greatly stabilize the dimer structure (Figure 3-12b).

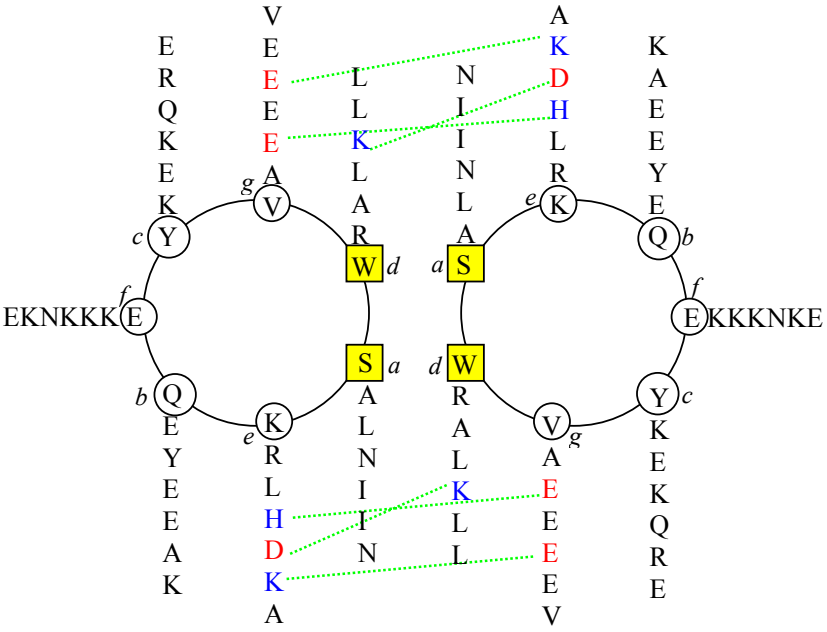
a

```
70      PESSENKNLG    GVTQESFDLM    IKENPSSQYW    KEVAEKRRKA
                                          abcd      efgabcdefg

110     LYEALKENEK    LHKEIEQKDN    EIARLKKENK    ELAEVAEHVQ
        abcdefgabc    defgabcdef    gabcdefgab    cdefg

150     YMA
```

b



c

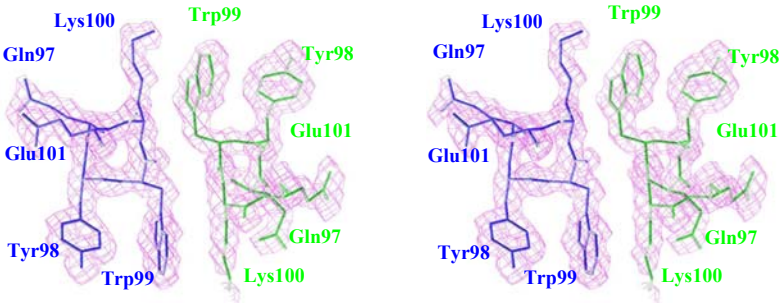


Figure 3-12. (a) The residues of the seven-heptad repeats of the Geminin coiled coil, assigned according to the crystal structure. Residues at the a and d positions are colored red. (b) A helical wheel representation of the repeated sequence of the Geminin coiled coil, highlighting the a and d positions at the center of the dimer interface. The positions of the heptad repeat are labelled a-g. Salt bridges between the two chains are represented by green dashed line. The positively charged residues are in blue while the negative charged residues are in red. (c) Stereo view of Trp99 located at the interface of the dimeric coiled coil. The $2F_o$-F_c electron density map is contoured at the 1.0σ level in O.

3.3.3 Surface of Geminin coiled coil

Although there are totally 16 positively charged residues and 14 negatively charged residues in the HGeminin92-152 sequence, the surface potential of the structure is mainly negative. In addition, near the N-terminus, Lys105, Arg106, Arg107 and Lys108 of the parallel coiled coil chains form a positively charged collar. The surface after this positively charged collar to the C-terminal end is highly negatively charged (Figure 3-13a, b). Such a distinctive charge surface of the Geminin coiled coil indicates that interaction between Geminin and Cdt1 may be based on strong electrostatic interaction.

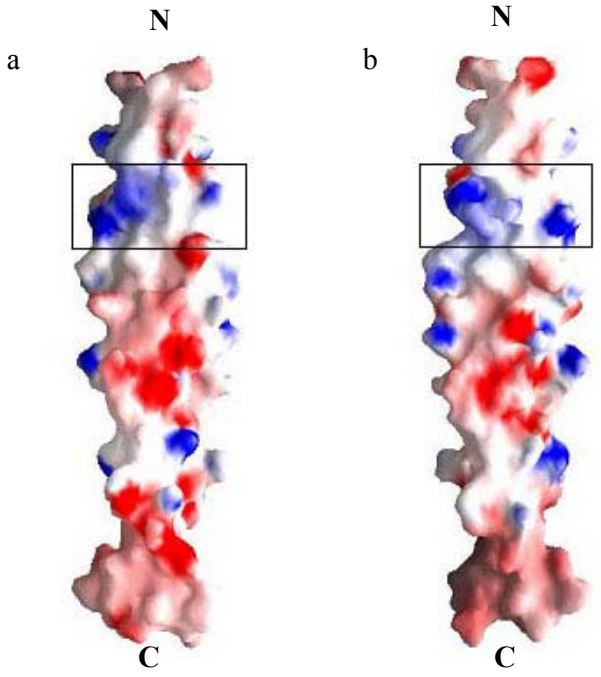


Figure 3-13. Electrostatic surface potential of Geminin92-152. (a) and (b) are the front and back views of one chain. The positively charged collar region is highlighted by the rectangular box. The dark red region indicates a potential of less than -10 kT/e, while dark blue indicates greater than 10 kT/e. The electrostatic potentials were calculated by GRASP (Nicholls, 1991).

3.4 DIMERIZATION OF GEMININ THROUGH ITS COILED COIL DOMAIN IS NECESSARY FOR ITS FUNCTION

3.4.1 Dimerization of Geminin through coiled coil region is necessary for its interaction with Cdt1 *in vitro* and *in vivo*

Based on the parallel coiled coil structure of HGeminin70-152, mutant Geminin LZ, in the Leucine Zipper residues that are located at the interface of the coiled coil, was generated by using the Quickchange site directed mutagenesis kit (Stratagene). In this mutant, Leu120, Ile124, Ile131 and Leu134 are changed to alanines. Geminin LZ can not form a dimer as wild type Geminin(Figure 3-14c). At the same time, it also fails to interact with Cdt1. However, with the point mutation of four glutamic acid residues (residues 116, 123, 130 and 137) to alanines, Geminin EtoA's still retain their ability to interact with Cdt1 in vitro (Figure 3-14a).

To demonstrate that these results were relevant *in vivo*, wild type Geminin and Geminin LZ were expressed as GST fusion derivatives by transient transfection of human 293T cells. An HA epitope tagged Cdt1 was expressed in the same cells by cotransfection. Purification of Geminin from these cells by affinity with glutathione sepharose beads revealed the preferential binding of HA tagged Cdt1 with wild type Geminin relative to Geminin LZ in cell extracts (Figure 3-14b). The GST epitope tag itself is known to form homodimers. Thus, the failure of GST-Geminin LZ to associate with Cdt1 indicates that the specific motif generated by the parallel coiled coil structure is necessary for Geminin to interact with Cdt1.

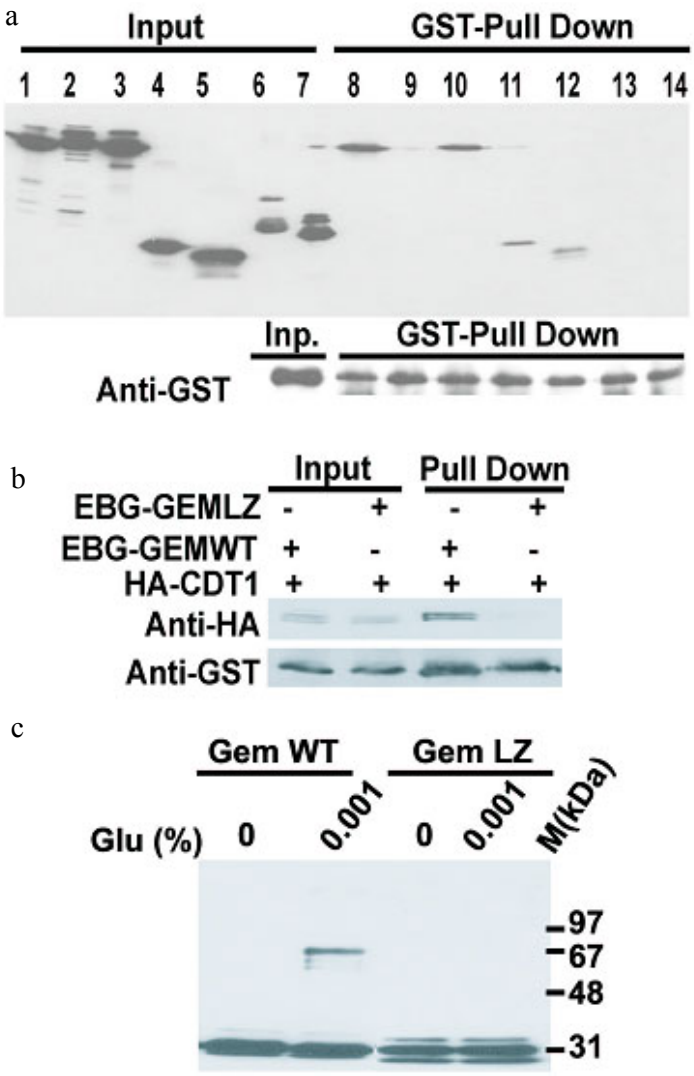


Figure 3-14. (a) Intact coiled coil domain of Geminin is required for Cdt1 interaction *in vitro*. Pull-down of His6-Geminin WT or mutants on GST-Cdt1 beads, followed by SDS-PAGE of bound proteins and immunoblot with anti-His antibodies. The input lanes are shown on the left. The forms of Geminin used were: WT (lanes 1, 8), LZ (lanes 2, 9), EtoA (lanes 3, 10), 70-152 (lanes 4, 11), 93-152 (lanes 5, 12), N114, (peptide after amino acid 114 is deleted, lanes 6, 13) and C115 (peptide before amino acid 115 is cut off, lanes 7, 14). The bottom panel shows

the presence of GST-Cdt1 in all the lanes. (b) *In vivo* interaction of Cdt1 with Geminin WT or Geminin LZ. 293T cells were transfected with HA-Cdt1 and GST-Geminin WT or GST-Geminin LZ. Input lanes show the expression of all the proteins whereas the pull down lanes show the proteins that are associated with glutathione agarose beads. Top: immunoblot with anti-HA and bottom: immunoblot with anti-GST. (c) Gluteraldehyde crosslinking of Geminin WT and Geminin LZ. 1 μg of each protein with equal amount of BSA was crosslinked by gluteraldehyde (concentrations at the top) and analyzed by 10% SDS-PAGE, followed by immunoblotting with anti His antibody.

3.4.2 Mutant Geminin LZ can not inhibit DNA replication in *Xenopus* egg extracts

Xenopus egg extracts can replicate sperm chromatin DNA when the two are incubated together. The incorporation of radioactive dNTP into sperm chromatin is followed by the electrophoresis of the products and autoradiography (Figure 3-15a). Quantitation of the replication products by Image-Quant analysis revealed that the addition of even 100 ng of wild type Geminin inhibited the replication reaction while 5 times as much Geminin LZ had no effect on the replication reaction (Figure 3-15b). Wild type Geminin associates with chromatin when it inhibits the replication reaction (Hodgson et al., 2002). Geminin LZ does not associate with the chromatin fraction (Figure3-15c), suggesting that the association of Geminin with chromatin might be mediated through its interaction with Cdt1 through its dimerization motif.

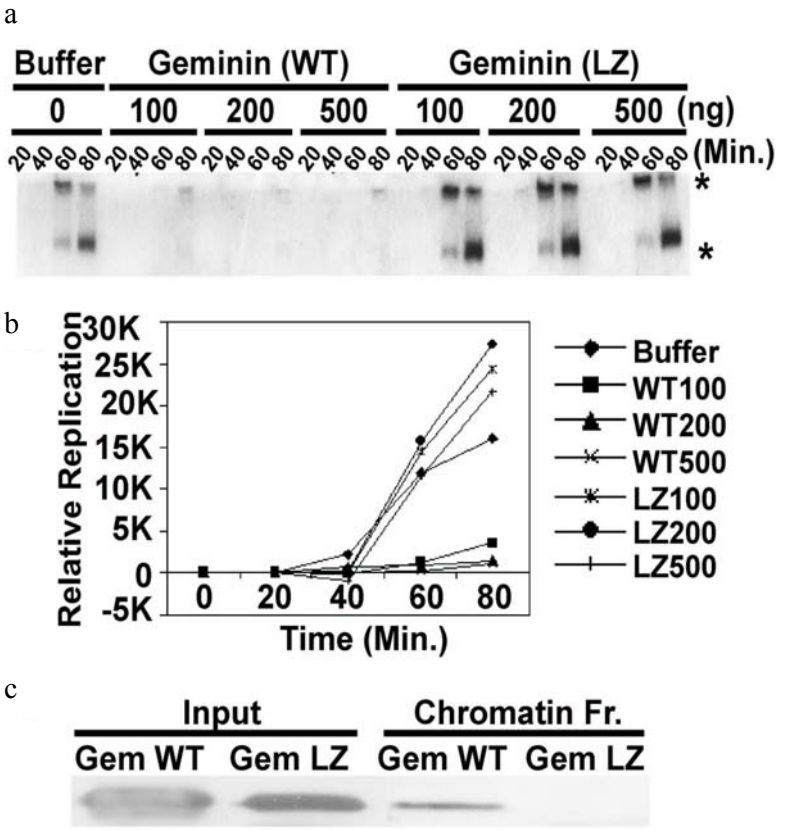


Figure 3-15. (a) Geminin LZ does not inhibit replication in *Xenopus* replication assay: *Xenopus* egg extracts were incubated with different amounts of Geminin WT or Geminin LZ, followed by the addition of sperm chromatin and α^{32}PdATP. The reaction mixtures were incubated at the room temperature before the addition of stop buffer at times indicated on top. Geminin WT inhibited the replication reaction at all concentrations whereas Geminin LZ did not have any effect on replication. "*" indicates the replication products. (b) Graphical representation of *Xenopus* replication assay: The bands corresponding to the replication products were analyzed by Phosphor imager and the relative intensities of the bands normalized using Image Quant

software. The relative units of replication are plotted against time. (c) Geminin LZ is not loaded on chromatin: *Xenopus* egg was extract incubated at room temperature for 40 min in the presence of 2 µg of each protein (His6-Geminin WT or LZ) and sperm chromatin and centrifuged through a sucrose cushion. The input mixtures and chromatin pellets were analyzed by SDS-PAGE and immunoblot with anti-His antibody.

3.4.3 Geminin LZ can not inhibit replication of EBV plasmid

The transient replication of episomes based on *ori*P of Epstein Barr virus in mammalian cells is dependent on components of pre-RC and inhibited by the over-expression of Geminin (Dhar et al., 2001). The same assay was performed with wild type Geminin and selected mutant derivatives (Figure. 3-16a). Cotransfection of the oriP based plasmid (p367) along with plasmid pEBG overexpressing wild type Geminin repressed the appearance of *Dpn*I resistant replicated plasmid. In contrast, Geminin LZ failed to suppress the replication of the p367 plasmid. Quantitation and normalization of replication data using phosphorimager showed that Geminin WT inhibited oriP replication to less than fifty percent while Geminin LZ had no effect (Figure 3-16b). Thus, mutant Geminin that fails to interact with Cdt1 also fails to suppress the replication of episomes in mammalian cells.

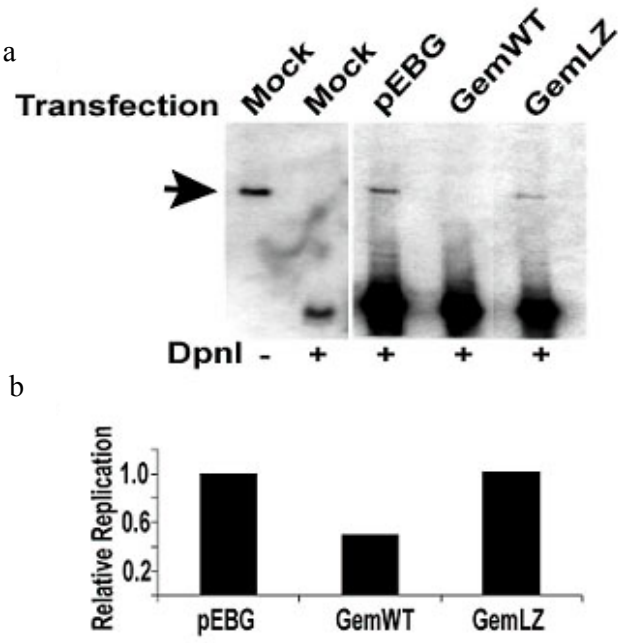


Figure 3-16. (a) Geminin LZ does not inhibit Epstein Barr Virus oriP mediated replication in a transient replication assay: Southern blot to detect DpnI resistant p367 DNA that has replicated in HCT116 cells following co-transfection of 2 μg of p367 and 4 μg of a plasmid expressing GST (lane 3), GST-Geminin WT (lane 4) or GST-Geminin-LZ (lane 5). Lanes 1 and 2 contain p367 mixed with mock transfected HCT116 DNA. (b) Quantitation of EBV replication assay: The bands corresponding to the *Dpn*I sensitive and *Dpn*I resistant digestion products were analyzed by Phosphor imager and the relative intensity of the resistant band was calculated for each lane using Image Quant software. The replication efficiency in each lane was plotted relative to the baseline in lane 3 (pEBG).

3.4.4 Geminin LZ fails to block the cell cycle

Transient overexpression of wild-type Geminin in cancer cells in culture did not produce a cell cycle block. However, a point mutation in the destruction box of Geminin (L26A), results in a protein that is stabilized in G1 and arrests cells at the G1-S transition (Shreeram et al., 2002; Wohlschlegel et al., 2002). The L26A and LZ mutations were combined to form Geminin L26A-LZ that is stable in G1. U2OS cells were co-transfected with a plasmid encoding farnesylated-GFP (Clontech) and any one of the following plasmids: pEBG-Geminin WT, pEBG-Geminin L26A, pEBG-Geminin LZ and pEBG-Geminin L26A-LZ. All the proteins were expressed following transfection as shown in Figure 3-17a. Cell cycle profiles of the GFP positive cells showed that pEBG-Geminin L26A blocked the cells in G1 but pEBG-Geminin L26A-LZ failed to do so (Figure 3-17b).

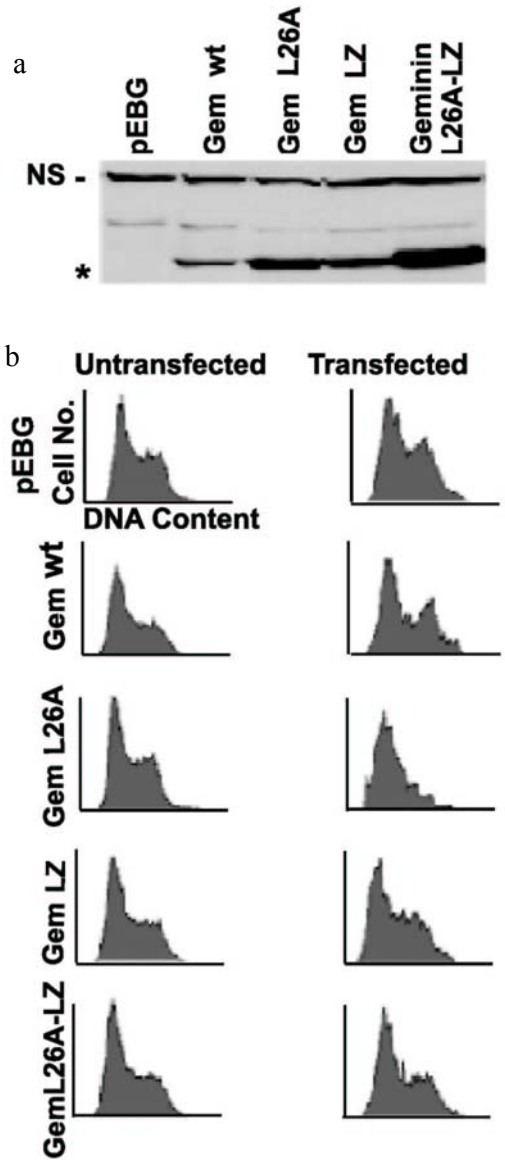


Figure 3-17. L26A-LZ Geminin does not block chromosomal replication. (a) Expression of GST Geminin in U2OS cells. Whole cell extracts prepared from U2OS cells transfected with pEBG (lane 1), pEBG Geminin wt (lane2), pEBG Geminin L26A (lane 3), pEBG Geminin LZ (lane 4) and pEBG Geminin L26A-LZ (lane 5) were immunoblotted with anti-GST antibody. "*" indicates the GST Geminin band and "NS" stands for he Non-specific cross-reactive band as a loading control. (b) The effect of overexpression of wild type and

mutant Geminin on cell cycle progression. U2OS cells were co-transfected with plasmids encoding farnesylated GFP and different forms of Geminin. The cell cycle profiles of GFP positive (transfected) and negative (untransfected) cells were determined 72 hours post transfection by propidium iodide staining and FACS analysis.

3.5 DISCUSSION

The crystal structure shows that Human Geminin 94-150 forms a parallel coiled coil dimer. Geminin is known to interact with Cdt1 and inhibit the DNA replication initiation process at exactly the step that requires Cdt1. Now, it is further understood that Geminin associates with itself through the coiled coil domain and the integrity of this coiled coil domain is essential for interaction with Cdt1. Point mutations in the leucine zipper (Geminin-LZ), which is critical in the hydrophobic interaction of the coiled coild structure, disrupts the dimerization and also abolishes its interaction with Cdt1. Point mutation of four glutamic acid residues (Geminin EtoA) retains the ability to interact with Cdt1 *in vitro*. Two deletion mutants (Geminin N114 and Geminin C115) that lack an intact coiled coil domain also fail to interact with Cdt1 (Figure 3-14a). Together, these observations suggest that the coiled coil domain of Geminin is essential for Cdt1 binding.

Besides being unable to form a dimer and interact with Cdt1, Geminin-LA also has lost the ability of inhibiting DNA replication in *Xenoupus* egg extract, EBV oriP based transient plasmid replication and cellular chromosomal replication. Combined with the evidence that monomeric Geminin fails to load on chromatin, it suggests that dimerization and interaction with Cdt1 is critical for the initial association of Geminin with chromatin and furthermore carry out Geminin's function. Since the depletion of Cdt1 also decreases the association of Geminin on chromatin (Gillespie, 2001), it is likely that the Geminin dimer interacts with Cdt1 to be loaded on chromatin where they interfere with Mcm2-7 loading, and thus disrupt the formation of the pre-replication complex.

Although the coiled coil domain of Geminin is necessary and sufficient for binding Cdt1, it is not sufficient to inhibit DNA replication. An additional region, residues 70-93, appears to be necessary for inhibiting DNA replication. The physical contiguity of residues 70-93 with the coiled coil domain might indicate that the critical function of this accessory domain may either stabilize the interaction with or make additional contacts with Cdt1 that interfere with whatever function is necessary for co-operating with Cdc6 to load the Mcm2-7 helicases. Alternatively, this domain of Geminin might have a novel interaction partner to help to carry out the work of dislodging the Mcm complex, which may be illustrated by conventional co-immunoprecipitation experiments. A targeted search of cellular proteins that are capable of interacting with this short portion of Geminin will help to distinguish between these possibilities. Overexpression of Geminin can selectively inhibit the replication of EBV based episomes while sparing the replication of cellular chromosomes (Dhar et al., 2001). It raised the possibility that a short peptide from Geminin could be developed for the purpose of eliminating EBV based episomes from EBV associated neoplasias where the virus and the viral oncogene is usually maintained without integration into the host chromosome. The fact that Geminin70-152 can inhibit EBV replication as full length Geminin shows that it is possible to design a peptide that can inhibit EBV replication, but such a peptide must have at least the coiled coil domain and an additional Cdt1 inhibitory domain of 23 residues (70-93) for its function. Further work on the Geminin70-152 and Cdt1 complex will definitely present more details.

PUBLICATIONS RELATED TO THIS THESIS

Yuan, P., Jedd, G., Kumaran, D., Swaminathan, S., Shio, H., Hewitt, D., Chua, N.H. and Swaminathan, K. A Hex1 crystal lattice required for Woronin body function in Neurospora crassa. *Nature Structural Biology,* **10**, 264–270 (2003). (PDB access code: 1KHI)

Yuan, P., Saxena, S., Dhar, S.K., Senga, T., Takeda, D., Robinson, H., Kornbluth, S., Dutta, A and Swaminathan, K. A dimerized coiled coil domain and an adjoining part of Geminin interact with two sites on Cdt1 for replication inhibition. (submitted). (PDB access code: 1UII)

APPENDIX A MEDIUM AND SOLUTION

1) Lysis buffer A

50 mM Tris (pH 7.0), 150 mM NaCl, 1 mM EDTA and 1 mM dithiothreitol (DTT)

2) Wash buffer B1

10 mM Tris (pH 7.5), 150 mM NaCl, 1% Triton X-100

3) Wash buffer B2

10 mM Tris (pH 7.5), 150 mM NaCl

4) Cleavage buffer C

50 mM Tris (pH 7.0), 150 mM NaCl, 1 mM EDTA, 1 mM DTT

5) M9 minimum medium

To 750 ml sterile deionized H_2O add

5X M9 salts	200 ml
20% glucose	20 ml
1M $MgSO_4$	2 ml

1M $CaCl_2$	100 ul
0.5% Thiamine	100 µl

add DH_2O to 1 L. Sterilize the solution with 0.22 µm filter (Corning)

5X M9 salts

Dissolve Na_2HPO_4 34 g, KH_2PO_4 15 g, NaCl 2.5 g, NH_4Cl 5 g in 1 liter DH_2O. Divide the solution to 200 ml per bottle. Autoclave for 15 min at 15 psi on liquid cycle.

6) TDE buffer

10 mM Tris (pH 7.5), 150 mM NaCl, 10 mM DTT, 1mM EDTA

7) TDET buffer

10 mM Tris (pH 7.5), 150 mM NaCl, 10 mM DTT, 1 mM EDTA, 1% Triton X-100

8) Lysis buffer D

50 mM Tris (pH 8.0), 150 mM NaCl, 0.01% NP40, 1mM PMSF

9) Wash buffer E

50mM Tris (pH 8.0), 150 mM NaCl, 20 mM imidazole, 1 mM PMSF

10) Elution buffer F

50 mM Tris (pH 8.0), 150 mM NaCl, 500 mM imidazole, 1 mM PMSF

11) Low salt solution

50 mM Tris (pH 8.0), 50 mM NaCl

12) High salt solution

50 mM Tris (pH 8.0), 1 M NaCl

13) Wash buffer G

50 mM Tris (pH 8.0), 150 mM NaCl, 1mM DTT, 1mMEDTA

14) lysis buffer H

50 mM Tris (pH 8.5), 500 mM NaCl, 10% glycerol, 1 mM PMSF

15) Wash buffer I

50 mM Tris (pH 8.5), 500 mM NaCl, 10% glycerol, 1 mM PMSF, 20 mM imidazole

16) Elution buffer J

50 mM Tris (pH 8.5), 500 mM NaCl, 10% glycerol, 1 mM PMSF, 500 mM imidazole

REFERENCES

Abrahams, J.P. and De Graaff, R.A. (1998) New developments in phase refinement. *Curr Opin Struct Biol.*, 8, 601-605.

Akey, D., Malashkevich, V. and Kim, P. (2001) Buried polar residues in coiled-coil interfaces. *Biochemistry*, 40, 6352-6360.

Armentrout, V.N. and Maxwell, D.P. (1974) Hexagonal inclusions in an ergosterol-free mutant of *Neurospora crassa*. *Can. J. Microbiol.*, 20, 1427-1428.

Baudhuin, P., Beaufay, H. and de Duve, C. (1965) *J. Cell. Biol.* 26, 219-243

Berbee, M.L.T., J.W. (ed.). (2001) *Fungal molecular evolution: gene trees and geologic time. in The Mycota VII part B Systematics and Evolution.* Springer-Verlag Berlin, Heidelberg.

Bradford M.M. (1976) A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248-254.

Brunger, A.T. (1998) Crystallography & NMR system: a new software suit for macromoecular structure determination. *Acta Crystallogr. D*, 54, 905-921.

Buerger, M.J. (1956) *Introduction to Crystal Geometry*. McGraw Hill, New York.

Buerger, M.J. (1959) *Vector space, and its application in crystal-structure investigation*. Wiley, New York.

Buerger, M.J. (1966) *X-ray Crystallography*. John Wiley & SONs, Inc, New York.

Buerger, M.J. (1990) *Historical atlas of crystallography*. Published for International Union of Crystallography by Kluwer Academic Publishers, Dordrecht; Boston.

Burkhard, P., Stetefeld, J., and Strelkov, S.V. (2001) Coiled coils: a highly versatile protein folding motif. *Trends cell Bio.*, 11, 82-88.

Chang, J.F., Hall, B.E., Tanny, J.C., Moazed, D., Filman, D., Ellenberger, T. (2003) Structure of the coiled-coil dimerization motif of Sir4 and its interaction with Sir3. *Structure*, 11, 634-647.

Chaudhuri, B., Xu, H., Todorov, I., Dutta, A., Yates, J.L. (2001) Human DNA replication initiation factors, ORC and MCM, associate with oriP of Epstein-Barr virus. *Proc Natl Acad Sci U S A.*, 98, 10085-10089.

Chung, S.I., Park, M.H., Folk, J.E. and Lewis, M.S. (1991) Eukaryotic initiation factor 5A: the molecular form of the hypusine-containing protein from human erythrocytes. *Biochim. Biophys. Acta*, 1076, 448-451.

Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W. and T., W.P. (1995) *Current Protocols in Protein Science*. John Wiley & Sons, Inc.

Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallography D*, 50, 760-763.

Collinge, A.J. and Markham, P. (1985) Woronin bodies rapidly plug septal pores of severed *penicillium chrystogenum* hyphae. *Exp. Mycol.*, 9, 80-85.

Crick, F.H.C. (1953) The packing of alpha-helices: simple coiled coils. *Acta Crystallography*, 6, 689-697.

Deacon, A.M., Weeks, C.M., Miller, R. and Ealick, S.E. (1998) The Shake-and-Bake structure determination of triclinic lysozyme. *Proc Natl Acad Sci U S A.*, 95, 9284-9289.

Dhar, S.K., Yoshida, K., Machida, Y., Khaira, P., Chaudhuri, B., Wohlschlegel, J.A., Leffak, M., Yates, J. and Dutta, A. (2001) Replication from oriP of Epstein-Barr virus requires human ORC and is inhibited by geminin. *Cell*, 106, 287-296.

Diffley, J.F.X. (2001) DNA replication: Building the perfect switch. *Current Biology*, 11, R367-R370.

Drenth, J. (1994) *Principles of Protein X-ray Crystallography*. Springer-Verlag New York, Inc., New York.

Elcock, A.H. and McCammon,.J.A. (2001) Identification of protein oligomerization states by analysis of interface conservation. *Proc Natl Acad Sci U S A.*, 98, 2990-2994.

Elfgang, C., Rosorius,O., Hofer, L., Jaksche, H., Hauber, J. and Bevec, D. (1999) Evidence for specific nucleocytoplasmic transport pathways used by leucine-rich nuclear export signals. *Proc Natl Acad Sci U S A*, 96, 6229-6234.

Esnouf, R.M. An extensively modified version of MOLSCRIPT that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 32–134 (1997).

Frazão, C., Sieker, L., Sheldrick, GM., Lamzin, VS., LeGall, J., Carrondo, MA. (1998) Ab initio structure solution of a dimeric cytochrome cv3 from *Desulfovibrio gigas* containing disulfide bridges. *Journal of Biological Inorganic Chemistry*, 4, 162-165.

Furey, W.S., S. (1997) PHASES-95: a program package for the processing and analysis of diffraction data from macromolecules. *Methods Enzymol.*, 277, 590–629.

Gillespie, P.J., Li, A., Blow, J.J. (2001) Reconstitution of licensed replication origins on Xenopus sperm nuclei using purified proteins. *BMC Biochem*, 2, 15.

Gould, S.J., Keller, G.A., Hosken, N., Wikinson, J. and Subramani, S. (1989) A conserved tripeptide sorts proteins to peroxisomes. *J. Cell Biol.*, 108, 1657-1664.

Hahn, T. (1998) *International Tables for Crystallography Voloume A Space-Group Symmetry*. Kluwer Acadedmic Publishers.

Head, J.B., Markham, P. and Poole, R.K. (1989) Woronin bodies from *Penicillium chrysogenum*: Isolation anf charaterization by analytical subcellular fractionation. *Exp. Mycol.*, 13, 203-211.

Hendrick, K. and Thornton, J.M. (1998) PQS: a protein quaternary structure file server. *Trends Biol. Sci*, 23, 358-361.

Hoch, H.C. and Maxwell, D.P. (1974) Protinaceous hexagonal inclusions in hyphae of *Whetzelinia sclerotiorum* abd *Neurospora crassa*. *Can. J. Microbiol.*, 20, 1029-1035.

Hodgson, B., Li, A., Tada, S. and Blow, J.J. (2002) Geminin becomes activated as an inhibitor of Cdt1/RLF-B following nuclear import. *Current Biology*, 12, 678-683.

Hofmann, W., Reichart, B., Ewald, A., Muller, E., Schmitt, I., Stauber, R.H., Lottspeich, F., Jockusch, B.M., Scheer, U., Hauber, J. and Dabauvalle MC. (2001) Cofactor requirements for nuclear export of Rev response element (RRE)- and constitutive transport element (CTE)-containing retroviral RNAs. An unexpected role for actin. *J Cell Biol.*, 152, 895-910.

Holm, L. and Sander, C. (1999) Protein folds and families: sequence and structure alignments. *Nucleic Acids Res.*, 27, 244-247.

Jedd, G. and Chua, N.H. (2000) A new self-assembled peroxisomal vesicle required for efficient resealing of the plasma membrane. *Nature Cell Biology*, 2, 226-231.

Jones, T.A., Zou, J.Y. Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, 47, 110-119.

Keller, G.A. (1991) Evolutionary conservation of a microbody targeting signal that targets proteins to peroxisomes, glyoxysomes, and glycosomes. *J. Cell Biol.*, 114, 893-904.

Kim, K.K., Hung, L.W., Yokota, H., Kim, R. and Kim, S.H. (1998) Crystal structures of eukaryotic translation initiation factor 5A from Methanococcus jannaschii at 1.8 Å resolution. *Proc. Natl. Acad. Sci. USA*, 95, 10419-10424.

Kroll, K.L., Salic, A.N., Evans, L.M. and Kirschner, M.W. (1998) Geminin, a

neuralizing molecule that demarcates the future neural plate at the onset of gastrulation. *Development*, 125, 3247-3258.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, 26, 283-291.

Lygerou, Z. and Nurse, P. (2000) Cell cycle - License withheld - Geminin blocks DNA replication. *Science*, 290, 2271-2273.

Markham, P. (1994) Occlusions of septal pores in filamentous fungi. *Mycol. Res*, 98, 1089-1106.

Markham, P. and Collinge, A.J. (1987) Woronin bodies of filamentous fungi. *FEMS Microbiology Letters*, 46, 1-11.

Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, 33, 491-497.

McGarry, T.J. (2002) Geminin deficiency causes a Chk1-dependent G2 arrest in Xenopus. *Molecular Biology of the Cell*, 13, 3662-3671.

McGarry, T.J. and Kirschner, M.W. (1998) Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell*, 93, 1043-1053.

McRee, D.E. (1999a) *Pratical Protein Crystallography*. Academic Press, San Diego.

McRee, D.E. (1999b) XtalView/Xfit - A Versatile Program for Manipulating Atomic Coordinates and Electron Density. *J. Structural Biology*, 125, 156-165.

Merritt, E.A. & Murphy, M.E.P. RASTER3D version 2.0 — a program for photorealistic molecular graphics. *Acta. Crystallogr. D* **50**, 869–873 (1994).

Miller, R., DeTitta, G.T., Jones, R., Langs, D.A., Weeks, C.M. and Hauptman, H.A. (1993) On the application of the minimal principle to solve unknown structures. *Science*, 259, 1430-1433.

Momany, M., Richardson, E.A., Van Sickle, C. and Jedd, G. (2002) Mapping Woronin body position in Aspergillus nidulans. *Mycologia*, 94, 260-266.

Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Gen.*, 11, 281-296.

Ohno, S. (1970) *Evolution by Gene Duplication*. Springer-Verlag, Berlin.

Olsen, L.J. (1998) The surprising complexity of peroxisome biogenesis. *Plant Molecular Biology*, 38, 163-189.

Orengo, C.A. et al. (1997) CATH — a hierarchic classification of protein domain structures. *Structure*, 5, 1093-1108.

O'shea, E.K., Klemm, J.D., Kim,P.S. and Alber, T. (1991) X-ray structure of the GCN4 leucine zipper, a two stranded, prallel coiled coil. *Science*, 254, 539-544.

Otwinowski, Z.M., W. (1997) Processing of X-ray diffraction data collected using oscillation mode. *Methods Enzymol.*, 276, 307-326.

Pearl, F.M. et al (2000) Assigning genomic sequences to CATH. *Nucleic Acids Res.*,

28, 277-282.

Peat, T.S., Newman, J.S., Waldo, G.S., Berenden, J. and Terwilliger, T.C. (1998) Structure of translation initiation factor 5A from Pyrobaculum aerophilum at 1.75 Å resolution. *Structure*, 6, 1207-1214.

Petersen, B.O., Lukas, J., Sorensen, C.S., Bartek, J. and Helin, K. (1999) Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. *EMBO J.*, 15, 396-410.

Quinn, L.M., Herr, A., McGarry, T.J. and Richardson, H. (2001) The Drosophila Geminin homolog: roles for Geminin in limiting DNA replication, in anaphase and in neurogenesis. *Genes & Development*, 15, 2741-2754.

Rhodes, G. (2000) *Crystallography Made Crystal Clear*. Academic Press, San Diego.

Rosorius, O., Reichart, B., Kratzer F., Heger, P., Dabauvalle, M.C. and Hauber, J. (1999) Nuclear pore localization and nucleocytoplasmic transport of eIF-5A: evidence for direct interaction with the export receptor CRM1. *J Cell Sci.*, 112, 2369-2380.

Saha, P., Chen, J., Thome, K.C., Lawlis, S.J., Hou, Z.H., Hendricks, M., Parvin, J.D. and Dutta, A. (1998) Human CDC6/Cdc18 associates with Orc1 and cyclin-cdk and is selectively eliminated from the nucleus at the onset of S phase. *Mol Cell Biol.*, 18, 2758-2767.

Schatz, O., Oft, M., Dascher, C., Schebesta, M., Rosorius, O., Jaksche, H., Dobrovnik, M., and Bevec, D. and Hauber, J. (1998) Interaction of the HIV-1 rev cofactor eukaryotic initiation factor 5A with ribosomal protein L5. *Proc Natl Acad Sci U S A*,

95, 1607-1612.

Schepers, A., Ritzi, M., Bousset, K., Kremmer, E., Yates, J.L., Harwood, J., Diffley, J.F. and Hammerschmidt, W. (2001) Human origin recognition complex binds to the region of the latent origin of DNA replication of Epstein-Barr virus. *EMBO J.*, 20, 4588-4602.

Schnier, J., Schwelberger, H.G., Smit-McBride, Z., Kang, H.A. and Hershey, J.W. (1991) Translation initiation factor 5A and its hypusine modification are essential for cell viability in the yeast Saccharomyces cerevisiae. *Mol. Cell. Bio.*, 11, 3105-3114.

Soundararajan, S., Jedd, G.J., Li, X., Chua, M.H. and Naqi, N.I. (2004) Woronin body function is essential for efficient pathogenesis by the rice-blast fungus *Magnaporthe grisea*. *Plant Cell*, in press

Shreeram, S., Sparks, A., Lane, D.P. and Blow, J.J. (2002) Cell type-specific responses of human cells to inhibition of replication licensing. *Oncogene*, 21, 6624-6632.

Stout, G.H. and Jensen, L.H. (1989) *X-ray Structure Determination*. John Wiley & Sons, Inc, New York.

Subramani, S. (1998) Components involved in peroxisome import, biogenesis, proliferation, turnover, and movement. *Physiol. Rev.*, 78, 171-188.

Swinkels, B., Gould, S., Bodnar, A., Rachubinski, R. and Subramani, S. (1991) A novel, cleavable peroxisomal targeting signal at the amino-terminus of the rat 3-ketoacyl-CoA thiolase. *EMBO J.*, 10, 3255-3262.

Tabak, H.F., Braakman, I. and Distel, B. (1999) Peroxisomes: simple in function but complex in maintenance. *Trends Cell Biol*, 9, 447-453.

Tada, S., Li, A., Maiorano, D., Mechali, M. and Blow, J.J. (2001) Repression of origin assembly in metaphase depends on inhibition of RLF-B/Cdt1 by geminin. *Nature Cell Biology*, 3, 107-113.

Taylor, T.N., Hass, H. and Kerp, H. (1999) The oldest fossil ascomycetes. *Nature*, 399, 648.

Tenney, K., Hunt, I., Sweigard, J., Pounder, J.I., McClain, C., Bowman, E.J. and Bowman, B.J. (2000) hex-1, a gene unique to filamentous fungi, encodes the major protein of the Woronin body and functions as a plug for septal pores. *Fungal Genetics and Biology*, 31, 205-217.

Terwilliger, T.C., Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D*, 55, 849-861.

Thepaut, M., Hoh, F., Dumas, C., Calas, B., Strub, M.P. and Padilla, A. (2002) Crystallization and preliminary x-ray crystallographic analysis of Human Geminin Coiled-coil domain. *Biochimica Et Biophysica Acta-Proteins and Proteomics*, 1599, 149-151.

Trinci, A.P. and Collinge, A.J. (1974) Occlusion of the septal pores of damaged hyphae of *Neurospora crassa* by hexagonal crystals. *Protoplasma*, 80, 56-67.

Uson, I., Sheldrick, G.M. (1999) Advances in direct methods for protein crystallography. *Curr Opin Struct Biol.*, 9, 643-648.

Valdar, W.S. and Thornton, .J.M. (2001) Conservation helps to identify biologically relevant crystal contacts. *J Mol Biol.*, 313, 399-416.

Wang. T.W., Lu, L., Wang D. and Thompson, J.E. (2001) Isolation and characterization of senescence-induced cDNAs encoding deoxyhypusine synthase and eucaryotic translation initiation factor 5A from tomato. *J Biol Chem*, 276, 17541-17549.

Wergin, W.P. (1973) Development of woronin bodies from microbodies in *Fusarium oxysporum f. sp. lycopersici. Proteoplasm*, 76, 249-260.

Wohlschlegel, J.A., Dwyer, B.T., Dhar, S.K., Cvetic, C., Walter, J.C. and Dutta, A. (2000) Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. *Science*, 290, 2309-2312.

Wohlschlegel, J.A., Kutok, J.L., Weng, A.P. and Dutta, A. (2002) Expression of geminin as a marker of cell proliferation in normal tissues and malignancies. *American Journal of Pathology*, 161, 267-273.

Woronin, M. (1864) Entwicklungsgeschichte der *Ascobolus pulcherrimus* Cr. und einiger Pezizen. *Abh.Senkenb. Naturforsch.*, 5, 333-344.

Xu, A., Chen, K.Y. (2001) Hypusine is required for a sequence-specific interaction of eukaryotic initiation factor 5A with postsystematic evolution of ligands by exponential enrichment RNA. *J Biol Chem.*, 276, 2555-2561.

Zuk, D. and Jacobson, A. (1998) A single amino acid substitution in yeast eIF-5A results in mRNA stabilization. *EMBO J.*, 17, 2914-2925.