©
כל הזכויות שמורות

אין להעתיק (במדיה כלשהי), להדפיס, לתרגם, לאחסן במאגר מידע, להפיץ באינטרנט, חיבור זה או כל חלק ממנו, למעט שימוש הוגן בקטעים קצרים מן החיבור למטרות לימוד, הוראה, ביקורת או מחקר.
שימוש מסחרי בחומר הכלול בחיבור זה אסור בהחלט.

©
All rights reserved

This work, in whole or in part, may not be copied (in any media), printed, translated, stored in a retrieval system, transmitted via the internet or other electronic means, except for "fair use" of brief quotations for academic instruction, criticism, or research purposes only.
Commercial use of this material is completely prohibited.

Structure Determination of Proteins Involved in the Stability of the Phycobilisomes During Enviromental Stress

Monica Dines

Structure Determination of Proteins Involved in the Stability of the Phycobilisomes During Enviromental Stress

Submitted in Partial Fulfillment of the
Requirements for the
Degree of Doctor of Philosophy

Monica Dines

Submited to the Senate of the Technion – Israel
Institute of Technology

Adar 5767 - Haifa - March 2007

This research was carried out in the department of chemistry under the supervisioin of Prof. Noam Adir

I wish to thank Prof. Noam Adir for his guidence, support, his enourmous patient and advice. He made me enjoy every day of our work together.

I wish to thank all my friends from Prof. Adir and Prof. Kaftory's laboratory.

Special thanks to Dr. Rakefet Schwarz who introduced me to the world of NblA.

Lastly, I would like to thank my wonderful family: Shlomi, Maya and Kineret.

The generous financial help of the Technion is gratefully acknowledged

List of Figures

Figure	Title	Page
1	Picture of cyanobacteria.	**4**
2	Schematic model of the PBS assembly process (Adir et al, 2006).	**6**
3	NblA proteins sequence comparation.	**7**
4	NblA *Anabaena* sp. PCC 7120 structure.	**7**
5	Representation of the proposed binding of NblA to phycocyanin (αβ) PC monomers (Bienert et al., 2006).	**8**
6	Schematic domain structure of different sHSPs	**9**
7	Comparision of the tertiary and quaternary structures of sHSPs.	**10**
8	Small heat shock proteins sequence comparison.	**11**
9	Model for the chaperon function of the sHSPs.	**14**
10	Crystallization techniques.	**20**
11	The expression of *Tv*-HspA in M15[pREP4] *E.coli* host.	**52**
12	SDS-PAGE of the purified proteins prior to crystallization.	**54**
13	A. Crystal of NblA from *S. elongatus* sp. PCC 7984 grown in the presence of 2M urea and ethylene glycol. The crystals dimensions are about 0.1x0.05x0.04mm. B. Crystals of NblA from *T. vulcanus* grown in presence of 2M urea and ethylene glycol, mounted in nylon loop. The crystal dimensions are about 0.06x0.04x0.01mm. C. *Tv*-sHSP crystal grown in the presence of 2M urea and PEG MME 550. The crystals dimensions are about 0.04x0.04x0.04mm.	**57**
14	a) 1° oscillation diffraction pattern to 2.5Å collected on beamline ID-29 at the ESRF, Grenoble. b) 1° oscillation diffraction pattern to 2.5Å of *Tv*-NblA collected on beamline ID-14-1 at the ESRF, Grenoble. c) 1° oscillation diffraction pattern to 2.2Å of *Se*-NblA collected on beamline ID-14-1 at the ESRF, Grenoble.	**59**
15	Theoretical location of K-edges for selenium.	**62**
16	X-ray fluorescence spectrum of the Se-Met-*Tv*-NblA crystal around the Se K- edge (1Å=12398.5 eV).	**63**
17	Homology based molecular model of NblA protein folded according to structure of Anabaena NblA .	**65**
18	The initial *Tv*-NblA structure obtained after the molecular replacement with Phaser.	**65**
19	Correlation between the *Tv*-NblA final model and the electron density map.	**66**
20	The organization of *Tv*-NblA subunits in the asymmetric unit.	**67**
21	The superposition of the chains of *Tv*-NblA from the same asymmetric unit.	**68**
22	Superposition of the *Tv*-NblA (green) and *Anabaena* NblA (blue).	**69**
23	Calculated electrostatic potential of *Tv*-NblA (upper) and *Anabaena* (lower).	**69**
24	Calculated electrostatic potential comparation between the models of NblA from different organisms from both sides.	**70**
25	Comparison between the α-helices pairs between the NblA and PC molecules.	**72**
26	Comparison between the electrostatic potential of the NblA and the helices pairs of the PC molecules.	**73**
27	Theoretical model of interaction between the NblA and the phycobilisome.	**75**

28	A) PC structure (1KTP) with the interactions between the α-helices pairs B) Interaction between the pair H and C.	**77**
29	Final structure of *Tv*-NblA without urea.	**79**
30	The superposition of the *Tv*-NblA without urea in (P1, green) and the *Tv*-NblA with urea (P3221, blue).	**80**
31	*Tv*-NblA without urea structure. The interaction between the different dimers is presented in red.	**80**
32	The initial *Se*-NblA model obtained after molecular replacement is an octamer.	**82**
33	Se-NblA model with an R-factor of about 40% is formed from three dimers.	**83**
34	A) *Se*-NblA preliminary model structure with the point mutations that prevent the bleaching process (red). B) *Se* –NblA dimer preliminary model with the point mutations that prevent the bleaching process and are related in the monomers interaction (red).	**84**
35	Homology based molecular model of the HspA protein folded according to the structure of wheat sHSP 16.9	**85**

List of Tables

Table	**Title**	**Page**
1	Native diffraction data sets statistics.	**60**
2	Summary of Se-Met-*Tv*-NblA crystal parameters and data collection statistics.	**63**
3	*Tv*-NblA reflection data and structure refinement statistics	**67**
4	Statistics of reflection data and structure refinement of the *Tv*-NblA with urea	**79**

List of Schemes

Scheme	**Title**	**Page**
1	Preparation of Se-NblA soluble protein for crystallization.	**51**
2	Preparation of soluble *Tv*-NblA and *Tv*-HspA for crystallization.	**52**

Table of Contents

1. Abstract 1
2. Introduction 2
2.1 Cyanobacteria 2
2.2 Photosynthesis and phycobilisome structure in cyanobacteria 3
2.3 Influence of the environment on the photosynthetic apparatus 4
2.4 Cyanobacterial response to nutrient stress 5
2.5 The structure and possible function of NblA 7
2.6 Cyanobacterial response to thermal stress 8
2.7 The structures of sHSPs 9
2.8 *In vivo* properties of sHSPs in other organisms 12
2.9 Regulation of sHSPs 12
2.10 The role of sHSPs in human disease 14
3. Research goals 16
4. Research plan and methodology 18
4.1 Protein isolation 18
4.2 Protein crystallization 18
4.2.1 Principles of protein crystallization 18
4.2.2 Crystallization techniques 19
4.3 X-ray diffraction 20
4.4 Data reduction 21
4.5 Structure determination 22
4.5.1 Molecular Replacement 23
4.5.2 Multiple Wavelength Anomalous Dispersion 24
4.5.3 Multiple Isomorphous Replacement 25
4.5.4 Calculation of electron density maps 25
4.5.5 Techniques for improvement of electron density maps 26
4.5.6 Map fitting 26
4.5.7 Model building 27
4.5.8 Model refinement 27
4.6 Functional analysis of the protein targets 30
5 Materials and Methods 32
5.1 Materials 32
5.1.1 General reagents 32
5.1.2 Biochemical, recombinant DNA technology and crystallization kits. 33
5.1.3 General materials 34
5.1.4 Purification chromatography columns 34
5.1.5 Growth media 34
5.1.6 Cell lines 35
5.1.7 Plasmid vectors and oligonucleotides 35
5.1.8 Vectors description 35
5.1.9 Oligonucleotides 35
5.1.10 Enzymes: 36
5.1.11 Buffers and solutions: 36
5.1.12 Column material and sources 37
5.1.13 Crystallization solutions 37
5.2 Methods 38
5.2.1 Molecular methods 38
5.2.2 Manipulation of plasmid DNA 39
5.2.3 Analytical methods 41

5.2.4 Overexpression of the recombinant NblA and HspA proteins. 41
5.2.5 Isolation, purification and crystallization of the proteins 42
5.2.6 Crystal handling and diffraction analysis 44
5.2.7 Computational software 45
6 Results and Discussion 49
6.1 Cloning of the *NblA* and *HspA* genes 49
6.2 NblA and HspA isolation and purification 50
6.3 NblA and HspA solubility problems.....55
6.3.1 Solving the HspA solubility problems 55
6.3.2 Addition of 2M urea solves the NblA and HspA solubility problem ..55
6.4 Crystallization of NblA.....56
6.5 Preliminary X-ray diffraction analysis and data set collection.....58
6.6 Data set analysis.....61
6.7 Structure determination of *Tv*-NblA 62
6.7.1 Selenium multiwavelength anomalous dispersion of Se-Met-*Tv*-NblA 62
6.7.2 Molecular replacement.....64
6.7.3 The overall structure of *Tv*-NblA.....67
6.7.4 NblA and phycocyanin interactions.....71
6.7.5 Model of NblA-Phycocyanin Interactions 73
6.7.6 Structure determination of *Tv*-NblA without urea 78
6.7.7 Structure determination of *Se*-NblA 81
6.7.8 Structure determination of HspA 84
7 Conclusions and plans for future research.....87

1. Abstract

The phycobilisome photosynthetic antenna complex, found in cyanobacteria and red-algae, interacts with proteins expressed specifically to deal with different forms of physiological stress. Under conditions of nutrient starvation, expression of the NblA protein is required in the process that leads to phycobilisome degradation and bleaching of the cells. HspA, a 16.5kDa heat shock protein expressed in cyanobacterial cells, has been shown to provide functional stability to the phycobilisome during heat stress. We have cloned the genes encoding for the NblA proteins from the thermophilic cyanobacteria *Thermosynecococcus vulcanus* (*Tv*-NblA) and from the mesophilic *Synecococcus elongatus* (*Se*-NblA) and for the HspA from *T. vulcanus* into bacterial expression vectors in order to determine their three-dimensional structures. The resulting recombinant proteins were found to be sparingly soluble, limiting their usefulness in the performance of crystallization experiments. We have developed a novel protocol that utilizes relatively high concentrations of urea to afford sufficient solubility to the protein. This has lead to the successful growth of diffraction quality crystals of these proteins. Complete data sets collected to 2-2.5Å from crystals of proteins show that the crystals are stable, and useful for structure determination.

We report here the three-dimensional structure of the *Tv*-NblA to 2.5Å with and without urea. The monomer has a helix-hairpin-helix motif which dimerizes into an open four-helical bundle in the asymmetric unit, identical to the NblA dimer from *Anabena*, whose structure was previously determined. Comparison between structures and homology based models indicate that while the NblA family is only slightly homologous, there are certain attributes of the protein that remain constant, and could be involved in its function that is probably a structure-related function. We proposed here, also, a new theoretical model of interaction between the NblA and the phycobilisome that presents a new possibility of insertion of the NblA into the whole phycobilisome.

The similarity between the structure of the *Tv*-NblA with and without urea shows that the presence of urea did not influence the protein structure, denaturation has not occurred and specific protein-protein interactions have been preserved. We believe that this protocol may be a generally advantageous method to obtain well diffracting crystals of sparingly soluble proteins.

2. Introduction

2.1 Cyanobacteria

One of the crucial requirements for most forms of life is oxygen, produced by sunlight and photosynthetic organisms. The ancestors of modern cyanobacteria (see figure 1) (previously called blue-green algae), appear to have evolved about three billion years ago and are considered to be the first photosynthesizing organisms capable of producing oxygen. It is generally accepted that the primitive cyanobacteria gave rise to plant chloroplasts through endosymbiotic events, thereby conferring the ability for photosynthesis to algae and plants (Asayama, 2006; Nowitzki et al., 1998) . Cyanobacteria are gram-negative eubacteria that serve as excellent model systems for the study of photosynthesis-related genes and proteins, due to the relative ease in their growth in culture and due to the existence of well characterized genetic and biochemical techniques.

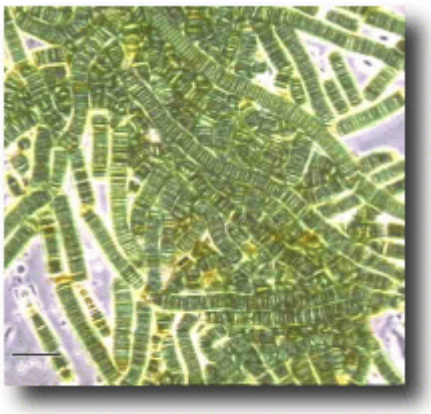


Figure 1. Picture of cyanobacteria.

Although they are true prokaryotes, cyanobacteria have a unique, elaborate and highly organized system of internal membranes which contain the photosynthetic apparatus. Chlorophyll a and several accessory pigments are embedded in these photosynthetic lamellae, the analogs of the eukaryotic thylakoid membranes.

2.2 Photosynthesis and Phycobilisome Structure in Cyanobacteria

The photosynthetic process starts by the absorption of light energy by pigments bound to large protein complexes called antennas. These antennas can transfer the absorbed energy with a very high efficiency to the reaction centres that perform the photochemical electron transfer reactions that convert light energy into chemical energy (Adir et al., 2006). The principal antenna complex present in prokaryotic (cyanobacteria) and some eukaryotes (red algae and cyanelles) is the phycobilisome (PBS). These are very large (3-7MDa) supramolecular protein complexes found on the stromal side of the thylakoid membrane. PBSs have a broad cross-section of absorption (500-680 nm) and transfer the absorbed energy in principal to photosystem II, but they can function, also, as antennas to photosystem I. All the PBSs are organized in a similar mode. They are formed from a basic building block called the monomer. The monomer is actually a heterodimer of two subunits (α and β) that are significantly homologous on the sequence and structural level. Their structures and mode of association lead to a very even and highly symmetric structure(Adir, 2005). The (αβ) monomers self-assemble in a number of apparently sequential steps:

1. three monomers associate into a trimeric $(\alpha\beta)_3$ disk with a diameter of about 110Ǻ, a width of 30Ǻ and an interior uneven aperture with a diameter 15-50 Ǻ;
2. two trimers associate into a hexameric $(\alpha\beta)_6$ double disk;
3. hexamers associate into one of two major structures, known as cores (containing allophycocyanin(APC)) and rods (containing phycocyanin(PC), and in some cases phycoerythrin(PE)).

The most common form of PBS complex is that composed of a tricylindrical core surrounded by six rods that form a hemidiscoidal semi-circle (Adir et al., 2006), (Glauser et al., 1992), (Glazer, 1989);(MacColl, 1998). An important remark is that while all the different $(\alpha\beta)_3$ disks are similar, the association of the disks to form core cylinders or rods is architecturally different than the forces that are involved in the rod-core assembly.

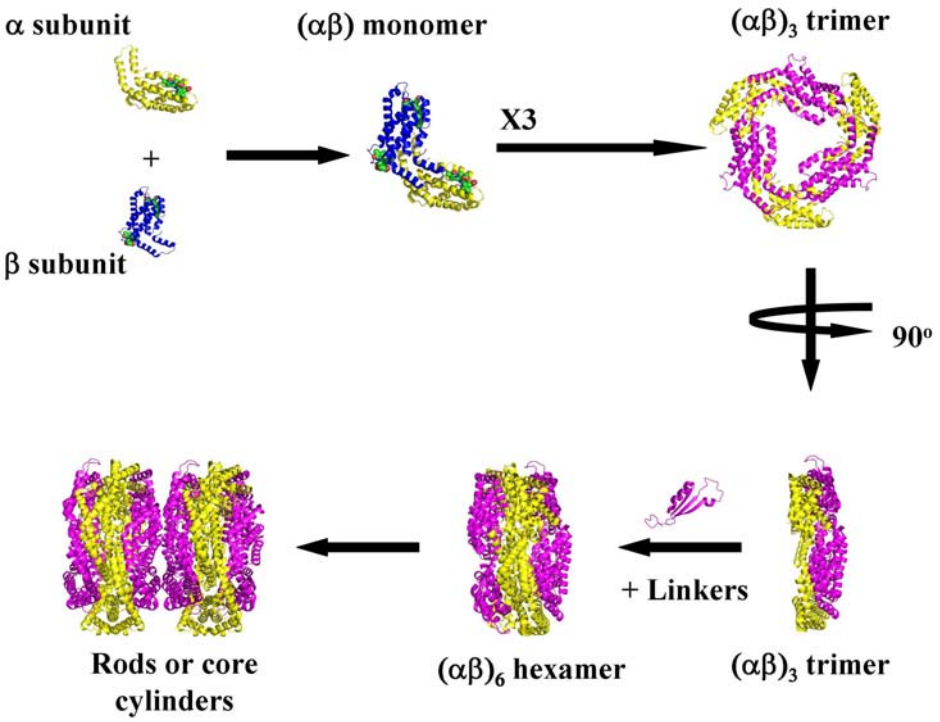


Figure 2. Schematic model of the PBS assembly process (Adir et al, 2006).

2.3 Environment influence on the photosynthetic apparatus

The activity of the photosynthetic apparatus is strongly influenced by the environment. Changes in the environmental parameters such as light quality, light quantity, temperature and water/nutrient availability require the activation of specific systems that perform important roles in the maintenance of the photosynthetic activity. Such maintenance systems typically entail changes in the levels of protein expression and can affect the pigment composition associated with photosynthetic complexes ((van Waasbergen et al., 2002); (Monson et al., 1992); (Grossman et al., 1995)).

The photosynthetic accommodation of stress is a dynamic process that operates on an intermediate timescale compared with short-term regulation and long-term developmental process. The scale and details of the stress response vary from species to species. In spite of this, a consistent response to high light is observed: increased levels of PSII, cytochrome *b/f* complex, ATP synthase, and components of the Calvin cycle; reductions in the levels of the major chlorophyll *a/b*-binding light

harvesting complexes associated with PSII. All these changes are reflected in increased capacities for oxygen evolution, electron transport and CO_2 consumption, and an increased ratio of chlorophyll a to b ((Anderson and Toole, 1998); Maenpaa and Anderson, 1989; Torre and Burkey, 1990; Bailey et al, 2001; Walters, 2004).
Water stress can initiate sustained decreases in the chlorophyll fluorescence indicative of increased thermal energy dissipation. This is part of the phenomenon of "photoinhibition''(Adams and Osmond, 1988). During salt stress, a decrease in the efficiency of photosynthetic energy conversion was also observed (Adams and Osmond, 1988). Nutrient stress, mainly as a result in nitrogen limitation, can cause a reduction of shoot growth and the capacity of photosynthesis and also increase the content of leaf zeaxanthin (Khamis et al., 1990).

2.4 Cyanobacteria response to nutrient stress

Cyanobacteria possess several mechanisms for modifying and/or protecting the composition of the photosynthetic machinery with respect to environmental stress. One of most notable examples of such a system is the response of cyanobacteria to the lack of the specific nutrients, especially nitrogen and sulfur. Under these conditions, the cyanobacterial cells change color from blue-green to yellow-green in a process known as chlorosis or bleaching (Allen and Smith, 1969). Using spectroscopy and other techniques it was found that the bleaching occurs within a few hours of the onset of nutrient starvation and was a result of the loss of the PBS in an ordered fashion (Grossman et al., 1993). It has been suggested that the disassembly of the PBS have two beneficial roles. During starvation the cellular metabolism decreases, while the rate of photosynthesis may be unchanged. This imbalance could lead to excessive absorption of excitation energy, which then leads to the production of harmful radical species(Adir and Lerner, 2003). Under extreme conditions this may lead to cell death and must be avoided. Since the rate of photosynthesis is directly coupled to the rate of energy absorption, uncoupling and degradation of the major antenna will be highly beneficial in avoiding overexcitation. A second role for PBS degradation would be to serve as an internal nutrient reservoir. The PBS can supply the cell with either amino acid residues for the synthesis of protein systems required for high affinity nutrient uptake, or by further degradation as building blocks for the synthesis of other

metabolites. Collier and Grossman used a genetic screen to identify mutants unable to degrade PBSs during nutrient starvation in order to identify the molecular pathway of PBS degradation (Collier and Grossman, 1994). The *NblA* (non-bleaching A) gene was thus identified and has been found in all cyanobacteria. Nitrogen is the major nutrient whose lacking can lead to a 50-fold increase in the amount of NblA protein (Collier and Grossman, 1994). In some cyanobacteria (such as *Synechococcus* sp. PCC 7942) sulfur and phosphorus limitation can also induce the *NblA* response (Collier and Grossman, 1994) while in other species sulfur limitation does not influence the expression of NblB (Richaud et al., 2001). During starvation, wild-type cells (but not *NblA* mutants) degrade two rod linker proteins: L_R 33 and L_R 34.5, indicating a possible functional connection between linkers and the NblA protein (Richaud et al., 2001). This observation supports previous data showing that wild-type cells degrade the PBS by first decreasing the length of the rods (Anderson and Toole, 1998); (Sidler, 1994). It was shown that in the cyanobacterium *Tolypothrix* PCC 7601, the NblA protein has affinity for both PC and PE subunits, but not for APC or for PBPs from other cyanobacterial species (Luque et al., 2003).

Beside the *NblA*, there are four other genes that have been identified as being related to PBS degradation: *nblR, nblS, nblB and nblC*. The NblR protein controls *NblA* expression and is critical for survival under stress conditions (Schwarz and Grossman, 1998). It has been suggested that *nblR* is controlled by *nblS*, a sensor histidine kinase that has a PAS domain (van Waasbergen et al., 2002). The NblB protein is necessary for PBS degradation during starvation, but unlike the NblA protein, the NblB levels are similar prior to or during nutrient starvation (Dolganov and Grossman, 1999). This protein is partially homologous with PCB lyases, the enzyme involved in chromophore attachment to the PC apoprotein. The most recent addition to the arsenal of proteins involved in the disassembly of the PBS is the NblC protein, which is required for *NblA* expression (Sendersky et al., 2005).

2.5 The Structure and Possible Function of NblA

The size of proteins belonging to the NblA family is quite small, ranging from 54–65 residues corresponding to molecular masses of about 6.5–7.0 kDa. The sequence identity between NblA proteins is relatively poor (about 30%) (see fig. 3), which is somewhat surprising since the PBS components, with which the NblA interacts are highly homologous (> 70%).

```
|nv|          MEQRFPELNVDLSFEQEFQMRVMEEQVSAMSLQEARELLLQASRLLMMKDNVIRSLVKR- 59
|se|          MLPPLPDF--SLSVEQQFDLQKYRQQVRDISREDLEDLFIEVVRQKMAHENIFKGMIRQ- 57
|Anna|         XNQPI-----ELSLEQQFSIRSFATQVQNXSHDQAKDFLVKLYEQXVVREATYQELLKHQ 55
                   :     .**.**:*.::    **   * :: .:::::  .  : ::   : ::::

|nv|          -----AARS- 63
|se|          -----GS--- 59
|Anna|         WGLDSGSTPA 65
```

Figure 3. NblA proteins sequence comparation.

In spite of the poor sequence identity the small NblA polypeptides always consists of two α-helixes connected by a small loop.

The structure of an NblA protein from *Anabaena* sp. PCC 7120 was recently determined (Bienert et al., 2006) (see fig 4). The structure also revealed that the monomer consists of a helix-hairpin-helix motif with a 37° angle between the two α-helixes, which dimerizes to form an open four-helical bundle. Although not stated specifically that the NblA dimers were present prior to crystallization in solution, it is implicitly stated that these authors propose the dimer as the basic structural unit of the NblA. The dimers are stabilized by only a very few polar contacts.

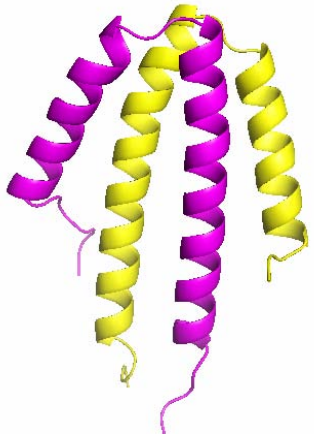


Figure 4. NblA *Anabaena* sp. PCC 7120 structure. The NblA dimmer with a helix-hairpin-helix motif

Using pull-down assays with immobilized NblA and peptide scanning, Bienert et al. proposed that the NblA interacts solely with the α-subunit of PE or PC. By site-directed mutagenesis they proposed that the NblA residues participating in this interaction are from the helical termini of the dimer (see fig. 5). This possible mode of interaction seems to us not very probable. Observation of the phycobilisomes models can show that the NblA molecule is too "large" to insert into the disk cavities of the phycobilisome. Thus it appears that many of the molecular details on how the NblA functions remain unclear.

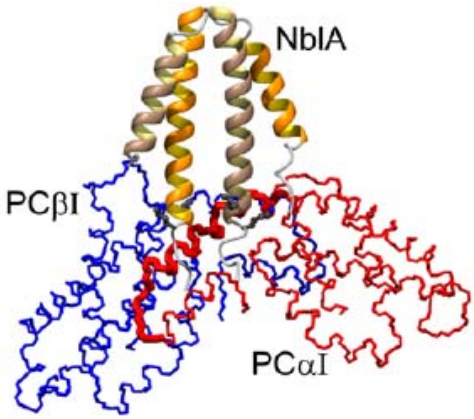


Figure 5. Representation of the proposed binding of NblA to phycocyanin (αβ) PC monomers (Bienert et al., 2006).

2.6 Cyanobacterial response to thermal stress

Another example of the response of cyanobateria to a change in its environment is the protection of photosystem II (PSII) and PC, the major constituent of the PBS, by a small heat shock protein (sHSP), one of the major proteins that accumulate upon heat-shock treatment in plants and cyanobacteria ((Roy et al., 1999; Roy and Nakamoto, 1998); (Nakamoto and Honma, 2006)). The sHSP in cyanobacteria also have a role in membrane stabilization during heat shock (Torok et al., 2001). The sHSP "acts like an amphitropic protein" during heat stress. It can bind weakly, reversibly, and specifically to membrane lipids. During heat stress many proteins are inactivated. The large reservoirs of sHSP that accumulate during the stress help the protein to reactivate after the stress has passed.

2.7 The structures of sHSPs

Most of the sHSPs investigated to date, share a conserved region of about 90 residues called the "α-crystallin-domain" in the C-terminal part of the protein. The α-crystalline-domain may be flanked by an N-terminal hydrophobic region, quite divergent in sequence and length and a short C-terminal extension. (Haslbeck, 2002) (see fig. 6). Because of their length and flexibility the N-terminal regions have only been partially determined in the structures already solved. In most of the small heat shock proteins the "α-crystallin-domain" amino acid sequence is variable. However, the structure of the compact β-sheet sandwich core is conserved throughout the sHSP family (see fig. 6). This core has an immunoglobulin-like fold with two layers of three and five antiparallel strands connected by a short interdomain loop. The "α-crystallin-domain" can dimerize to form an intersubunit composite β-sheet. This dimerization appears in most of the sHSPs whose structures have been determined.

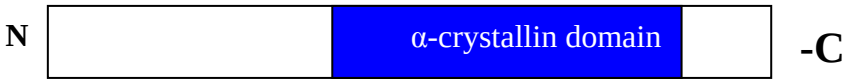


Figure 6. Schematic domain structure of different sHSPs

One of the common characteristics of all the sHSPs is their organization in large oligomeric structures, comprising 9-50 subunits. The structures of seven members of the sHSPs family have been determined (three by cryoelectron microscopy, and four by X-ray diffraction) (see fig. 7), most of them revealing hollow, globule-like structures. The oligomers formed by the α-crystallin domains show a variable quaternary structure (Haley et al., 1998a), but on the other hand, the quaternary structures of HSP 16.5 from Methanococcus jannaschii (Kim et al., 1998) and of wheat HSP 16.9 (van Montfort et al., 2001) are well defined. HSP 16.5 from Methanococcus jannaschii forms a 24-subunit complex, in contrast with wheat HSP 16.9 which assembles into a dodecameric double disk (see fig. 7). Each disk is organized as a trimer of dimers.

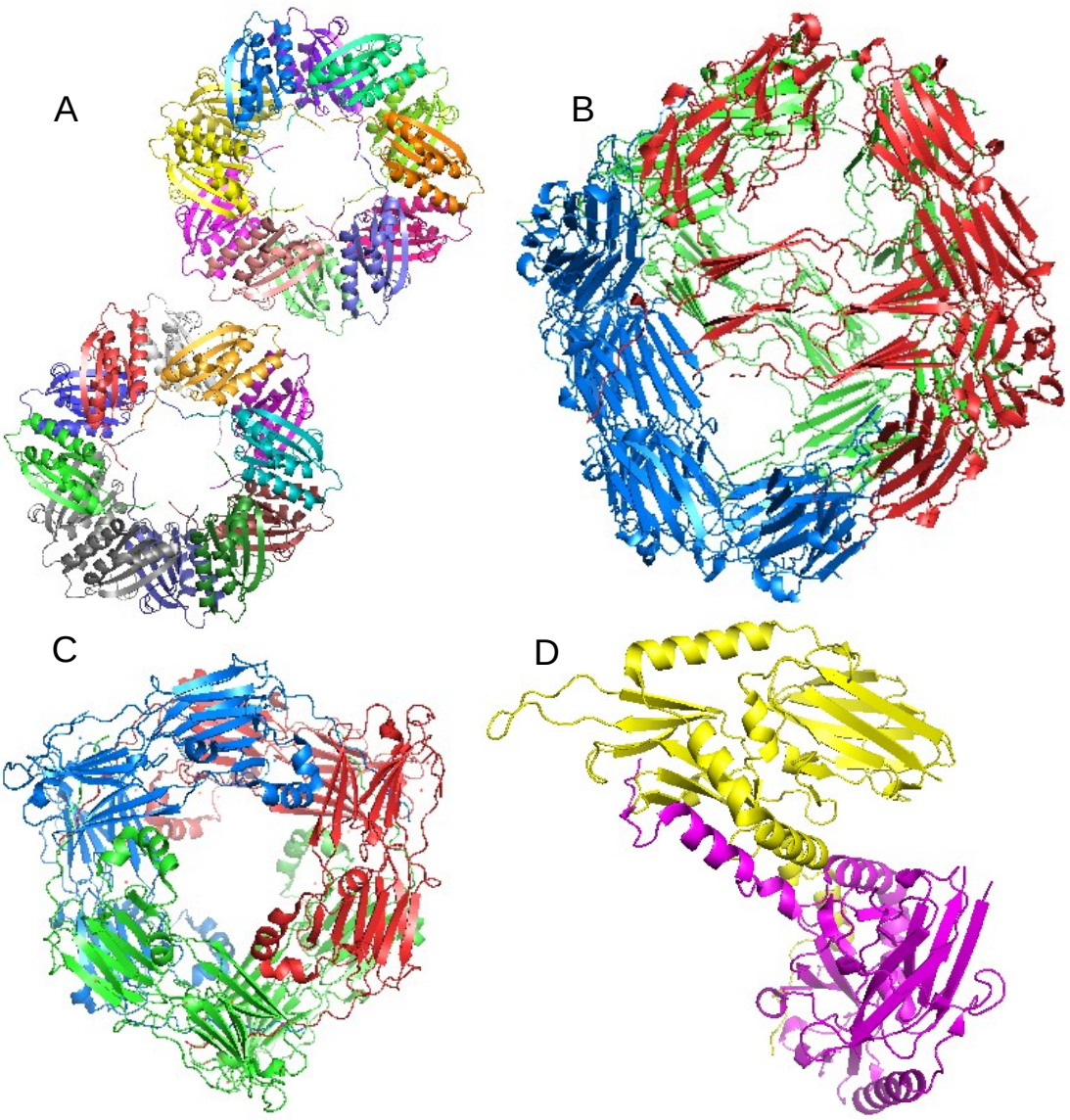


Figure 7. Comparision of the tertiary and quaternary structures of sHSPs. (A) Crystal structure of SP1 from *Populus Termula* (Dgany et al., 2004). (B) Crystal structure of sHSP16.5 from *M. jannaschii* (Kim et al., 1998). (C) Crystal structure of wheat sHSP16.9 (van Montfort et al., 2001). (D) Crystal structure of a relatively large sHSP (monomer) from a metazoan TSP36 (Stamler et al., 2005)

In both the HSP16.5 and 16.9 the N-terminal region is localized in the interior of the oligomeric structures. In the HSP16.5 and the α-crystallin oligomer, these unordered regions sequester inside the sphere, while in the case of HSP 16.9 they are buried in the interior of the oligomeric structures (Haslbeck, 2002) (see fig. 7). Recently the structure of a monodisperse metazoan sHsp, Tsp36 was determined (Stamler et al., 2005). In the case of TSP36 the α-crystallin domain structure is similar to that of HSP16.5 and 16.9 and it is also formed from a β-sandwich sub-structure

with a seven strand arrangement. All of the α-crystallin domains contain a large, flexible loop (figure 7). Unlike the other sHSPs, the TSP36 form only a dimer in its crystal form.

As can be observed in figure 8 no amino acid sequence similarity exists between SP1 and the other sHSPs. Radical differences also were found in the crystal structure of the SP1 and the other three sHSPs whose structures were previously solved (figure 8), indicating that SP1 does not belong to the conserved sHSPs family (Dgany et al., 2004). This may suggest that SP1 may represent a different family of proteins with a different and unique role in the thermostability.

```
HspA            --------------------------------------------------MALVR----W 6
Eukaryotic      --------------------------------------------------MSIVRRSNVF 10
MJ              -------------------------------------------------MFGRDPFD 8
Tsp36           MSIFPTRDSRDLSSRRRSLIDWEFPQMALVPLDQVFDWAERSRQSLHDDIVNMHRNLFSL 60
                                                                  :

HspA            EPFREIDAIQRQMNRLFDELIPL-TERRSDLSFLPAAELEETPEALLLKVELPGMDPKDI 65
Eukaryotic      DPFADLWADPFDTFRSIVPAISG-GGSETAAFANARMDWKETPEAHVFKADLPGVKKEEV 69
MJ              SLFERMFKEFFATPMTGTTMIQSSTGIQISGKGFMPISIIEGDQHIKVIAWLPGVNKEDI 68
Tsp36           EPFTAMDNAFESVMKEMSAIQPR--EFHPELEYTQPGELDFLKDAYEVGKDGRLHFKVYF 118
                . *  :                      .         .     :   .           .

HspA            DVQVT-AEAVSISGERKSETKTETEGMKRTEF-RYGKFQRVIPLPVRIQNTSVKAEYKDG 123
Eukaryotic      KVEVEDGNVLVVSGERTKEKEDKNDKWHRVER-SSGKFVRRFRLLEDAKVEEVKAGLENG 128
MJ              ILNAV-GDTLEIRAKRSPLMITESERIIYSEIPEEEEIYRTIKLPATVKEENASAKFENG 127
Tsp36           NVKNFKAEEITIKADKNKLVVRAQKSVACGDAAMSESVGRSIPLPPSVDRNHIQATITTD 178
                 ::   .: : : ..:.       .     :     .. * : *    .    .*    .

HspA            ILHLTLPKAEEEKNRVVKVSLA------------------------------------ 145
Eukaryotic      VLTVTVPKAEVKKPEVKAIQISG----------------------------------- 151
MJ              VLSVILPKAESSIKKGINIE-------------------------------------- 147
Tsp36           DVLVIEAPVNEPNYKAIKLSPEKGLAIQPSEVQERQLAVKNKEGLEIVTAEDGSKKIHLE 238
                 : :  . .:    .   :.

HspA            ------------------------------------------------------------
Eukaryotic      ------------------------------------------------------------
MJ              ------------------------------------------------------------
Tsp36           LKVDPHFAPKDVKVWAKGNKVYVHGVTGKEEKTENASHSEHREFYKAFVTPEVVDASKTQ 298


HspA            ----------------
Eukaryotic      ----------------
MJ              ----------------
Tsp36           AEIVDGLMVVEAPLFK 314
```

Figure 8. Small heat shock proteins sequence comparison. HspA - small heat shock protein from *T. vulcanus*. Eukaryotic - wheat small heat shock protein 16.9. MJ – small heat shock protein from *M. jannaschii*. Tsp36 – small heat shock protein from metazoan

2.8 *In vivo* properties of sHSPs in other organisms

Under physiological conditions the abundance of sHSPs varies from organism to organism and is dependent on the type, growth cycle, differentiation and the oncogenic status of the cell (Arrigo, 2000; Bond and Schlesinger, 1986).

It has been observed that some prokaryotic sHSPs are tightly associated with intracellular aggregated endogenous proteins (Laskowska et al., 1996) and with inclusion bodies, which are formed during high level protein expression (Haslbeck, 2002). These results may suggest that the biological function of sHSPs is involvment in prevention of intracellular protein aggregation and the may have a chaperone activity. It was also observed that sHSPs could interact with different components of the cytoskeleton, like actin (Benndorf et al., 1994; Lavoie et al., 1993; Zhu et al., 1994) or intermediary filaments (Nicholl and Quinlan, 1994). These observations can be interpreted as showing that sHSPs, aside from their role as molecular chaperons that stabilize thermolabile intracellular proteins, may also have a role in preserving cell shape and cell integrity under stress conditions.

The overexpression of sHSPs has been observed to confer thermoresistance in a number of organisms and cell types (Nicholl and Quinlan, 1994),(Jofre et al., 2003; Knauf et al., 1992; Rollet et al., 1992). In contrast, yeast sHSP26 does not seem to have a demonstrable function in thermoprotection or during development (Petko and Lindquist, 1986; Susek and Lindquist, 1989). It was also observed that the deletion of sHSPs in *E. coli* is not lethal and does not have a recognizable phenotype (Thomas and Baneyx, 1998).

2.9 Regulation of sHSPs

Cofactors or ATP regulates many chaperons. In the case of sHSPs the role of ATP in functional regulation it is still disputed. It was observed that ATP enhanced molecular chaperone function of the sHSP human αB-crystallin by structural rearrangements and exposure of hydrophobic surfaces (Muchowski and Clark, 1998; Muchowski et al., 1999). In contrast, the chaperon activity of sHSPs *in vitro* has so far been found to be independent of ATP binding and hydrolysis . Under physiological conditions, sHSPs organized into high oligomeric complexes, which are proposed to

be an inactive storage form (Haslbeck, 2002). Elevated temperatures lead to the dissociation of these complexes into smaller oligomers (most of them dimers). In the case of yeast HSP26, the dimers specifically recognized and bound nonnative polypeptide chains, resulting in the cooperative formation of large HSP26-substrate complexes (Haslbeck et al., 1999). Once formed, the complexes between the sHSPs and substrates are very stable at physiological temperatures (Haslbeck et al., 2005).
An additional factor that seems to be associated with the changes in the oligomeric structure of sHSPs complexes is phosphorylation. Phosphorylation has been observed to be necessary, but not sufficient for dissociation of mammalian sHSP oligomers(Haslbeck, 2002).

Several sites in the N-terminal and also in the C-terminal region were observed to be especially involved in the chaperone function and complex formation of the sHSPs. These sites are in general hydrophobic (Basha et al., 2004; Stromer et al., 2004). For example in wheat sHSP16.9 the hydrophobic N-terminal region is related with oligomeric assembly (van Montfort et al., 2001). In yeast sHSP26, deletion of the corresponding hydrophobic N-terminal region permits dimerization, but with out chaperone activity.

Until now only a few members of the diverse family of sHSPs have been investigated in detail, and the exact mode of how sHSPs work and what regulate them is still unknown.

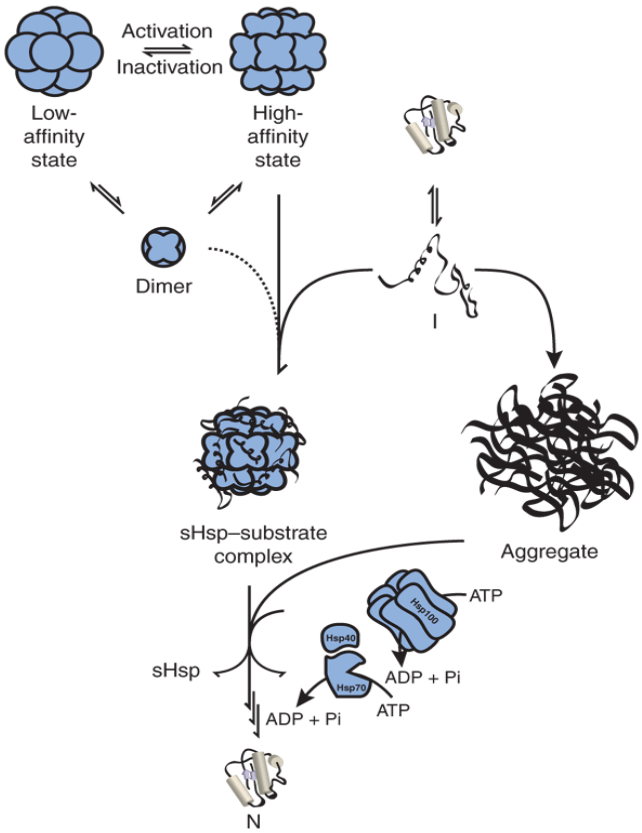


Figure 9. Model for the chaperon function of the sHSPs (Haslbeck et al., 2005).

2.10 The role of sHSPs in human disease

The list of human diseases involving protein aggregation continues to grow. To date, aggregation has been associated with sickle cell disease (Perutz et al., 2002), prion disease (Radford et al., 1999), cystic fibrosis (Johnston et al., 1998), Mallory bodies of hepatic cirrhosis (Irie et al., 1991), desmin-related myopathy (DRM) (Perng et al., 1999), cataract formation (Litt et al., 1998) and neurodegenerative diseases, including Parkinson's disease, diffuse Lewy body disease, Huntington's disease, Creutzfeldt–Jakob disease, Alzheimer's disease and Alexander's disease, in which sHSPs are a major component of Rosenthal fibers, (Renkawek et al., 1994), (Wisniewski and Goldman, 1998). The significant feature of these diseases is the protein aggregation that results from partial unfolding and exposure of hydrophobic surfaces, thus increasing attractive forces among protein molecules (Benedek et al., 1999; Canet et al., 1999; Xiao et al., 1999). Aggregation is thought to be inhibited when sHSPs are upregulated and form transient interactions with exposed

hydrophobic surfaces on unfolding proteins. There is no direct experimental support for the ‘chaperoned’ refolding of disordered polypeptides by sHSPs, although interactions between sHSPs and unfolding proteins are well documented(Haley et al., 1998b), (Ehrnsperger et al., 1997). Several studies indicate that binding between sHSPs and target substrate proteins involves ‘molten globule’ intermediates and complex kinetics (Canet et al., 1999);(Ehrnsperger et al., 1997). Although progress is being made with regard to understanding protein folding and unfolding, the mechanisms by which aggregates can kill cells or cause the cellular pathology of unfolding/misfolding diseases remains to be established. In cataracts, however, protein aggregates are known to be responsible for both lens opacity and visual impairment (Benedek et al., 1999). Until the direct involvement of sHSPs in protein folding or protective binding in aggregation diseases is demonstrated *in vivo*, the function of sHSPs in protection against protein aggregation in human disease remains theoretical.

3. Research goals

The photosynthetic apparatus undergoes substantial modulations in response to changing environmental parameters, such as light quality, light quantity, temperature and water/nutrient availability. Such systems typically entail changes in the levels of protein expression and can affect the pigment composition associated with photosynthetic complexes (Adir et al., 2003; Grossman et al., 1995; Kanervo et al., 2005; Monson et al., 1992; van Waasbergen et al., 2002).

Cyanobacteria possess several mechanisms for modifying and/or protecting the composition of the photosynthetic machinery with respect to environmental stress. One example of such a system is the response of cyanobacteria to the lack of the specific nutrients, especially nitrogen and sulfur. Under these conditions, occur the degradation of the light-harvesting pigment protein complex, the phycobilisome (Grossman et al., 1993). A principal protein component that is involved in the PBS degradation during nutrient deprivation NblA (Collier and Grossman, 1994).

Another example of the response of cyanobateria to a change in its environment is the protection of PSII and phycocyanin by HspA in the case of heat hock (Nakamoto and Honma, 2006; Roy et al., 1999; Roy and Nakamoto, 1998). Since structures of PBS components have been determined (Adir, 2005; Adir and Lerner, 2003; Adir et al., 2002; Brejc et al., 1995), high-resolution structures of the NblA and HspA proteins may provide information on their mode of action on the PBS.

One of the major bottlenecks in obtaining high-resolution structural data by X-ray crystallography is the lack of solubility of many recombinant proteins (Pusey et al., 2005; Yokoyama, 2003). The problem of recombinant protein solubility has become more apparent with the advent of high-throughput structural proteomics efforts, in which targets with solubility problems are many times discarded.

The specific goals of the research are the following:

1. Isolate, purify and characterize functional recombinant NblA and HspA protein of a quality that will afford the growth of crystals that will enable structure determination.
2. Determine the three-dimensional structures of the recombinant NblA and HspA proteins by X-ray crystallography methodologies.

3. Analyze the crystal structures in terms of the structural similarity with other members of the NblA and HspA family and propose potential mechanisms of action and interaction with the PBS and PSII.

4. Research plan and methodology

The work plan is built from the following steps:

4.1 Protein isolation

To determine the 3D structure by X-ray diffraction we need a relatively large amount of very highly purified protein (more than 99%). The minimum amount of protein necessary for the initial crystallization trials is usually around 5-10 mg. To obtain crystals, the protein molecules should usually be both chemically and structurally homogeneous (Drenth, 1994).

The procedures used in the purification of proteins are based on exploring the (Hendrickson et al., 1988) differences in physico-chemical properties of the macromolecules to be separated. The most commonly used techniques in the determination of sample purity are electrophoretic analyses. SDS-PAGE gives information on the degree of chemical purity and an estimate of the molecular weight of the protein subunits, but nothing else can be concluded regarding the structural integrity of the molecule. Therefore other methodologies should be employed such as HPLC, dynamic light scattering, etc.

4.2 Protein crystallization

4.2.1 Principles of protein crystallization

Identification of the conditions in which a protein crystallizes is an empirical process that involves a large number of crystallization trials in different conditions. “Beginner’s luck is definitely a factor in finding conditions for crystallizing a protein for the first time.” (McRee, 1999). The process of protein crystallization begins when supersaturated solutions of the protein of interest are produced. This leads to the induction of the formation of small aggregates, which serve as nuclei for the protein crystal.

In order to achieve super-saturation, precipitating agents are frequently included in the crystallization medium. In general, proteins are crystallized from

aqueous solutions with a concentration ranging from 1-50 mg/ml. The precipitation of the protein can be induced by increasing the effective concentration of the protein by the addition of precipitating agents such as salts (for example, ammonium sulfate) which will promote salting out or by the addition of organic molecules such as ethylene glycol or polyethylene glycol.

The number of variables involved in the crystallization process can be very large. Some of the principle factors which will determine the success of a given trial include the purity and concentration of the protein sample, temperature, pH, the presence and type of precipitating agent and additives (anions, cations, alcohols, detergents etc.).

4.2.2 Crystallization techniques

There are several techniques available for the crystallization of macromolecules including sitting drop vapor diffusion, hanging drop vapor diffusion, sandwich drop, microbatch under oil, microdialysis, and free interface diffusion (Rhodes, 1993). The most frequently used technique is vapor diffusion. The difference in concentration between the drop and the reservoir drives the system toward equilibrium by diffusion through the vapor phase. Free interface diffusion is one of the methods used by NASA in microgravity crystallization trials. Using this method the sample is in liquid contact with the precipitant and over time the sample and the precipitant diffuse into one another and crystallization may occur at the interface (Smatanova, 2002). Microbach under oil crystallization is a method where a small drop (0.2 μl) of the sample combined with the crystallization agent is pipetted under a layer of oil that allows water vapor to permeate from the drop. As a result the sample and the reagent concentrate. Unless the drop is equilibrated with a reservoir, water will leave the drop until only crystals remain (Smatanova, 2002) (see figure 10).

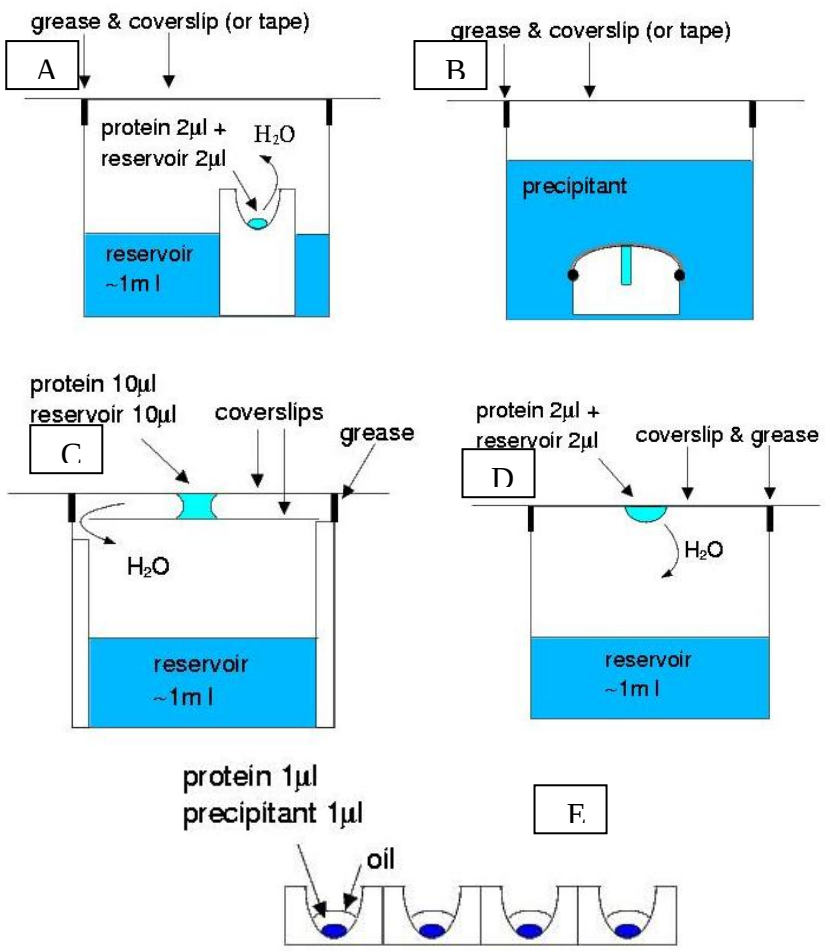


Figure 10. Crystallization techniques. (A) Sitting drop. (B) Microdialysis (C) Sandwich drop (D) Hanging drop (E) Microbatch under oil

4.3 X-rays diffraction

A protein crystal is a regular repeat of molecules, usually with some sort of internal rotational symmetry. Protein crystals are usually about 30-60% solvent by weight and are thus fragile and sensitive to drying out. The basic smallest repeating unit that can generate the crystal with only translational operations is termed the unit cell and is defined by three non-coplanar vectors a, b and c.

The unit cell itself can posses internal symmetry in which two or more identical chemical objects from within the unit cell are related by a symmetry element (a rotation axis, center of symmetry, screw axis, glide planes or mirror plane). The asymmetric unit represents the minimum structure necessary to generate the whole crystal using the elements of crystallographic symmetry. In protein crystallography

the asymmetric unit can correspond to a single molecule of the protein, but it may be composed of two or more molecules related to each other by symmetry axes, which is not crystallographic (i.e. they do not apply to all molecules of the crystal but rather are local in nature in that they apply to only a given asymmetric unit). Such axes are called non-crystallographic and occur frequently in protein crystals.

Each diffracted beam is defined by three parameters: its amplitude, its wavelength (which is fixed by the experimenter) and is phase angle. The amplitude and phase of a given reflection (h) are described by a quantity termed the structure factor F(h), whose amplitude is proportional to the square root of the intensity of the diffraction spot, which can be readily measured. The phase component of the structure factor is determined by position of a wave's maximum relative to an origin and can not be experimentally measured. Knowledge of the phase component is an essential requirement for structure determination.

4.4 Data reduction

The three-dimensional crystal lattice gives rise to a three-dimensional lattice of diffraction spots called a reciprocal lattice. The reciprocal lattice describes where on the detector the diffraction spots will occur. The individual reflections of the reciprocal lattice are designated Miller indices and termed h, k, l which are integers (either positive or negative). The Miller indices give the number of lattice spacing from the origin of the reciprocal lattice to the reflection in the directions of the vectors a, b and c.

After the collection of the diffraction data, the reflections are indexed (their Miller indices are determined) and their intensity placed on the same scale. An evaluation of the quality of the processed data can be made using several objective criteria. Resolution is a loosely used term that referees to the minimum interplanar spacing of the real lattice for the corresponding reciprocal lattice point (reflection) that is being measured. It is directly related to the optical definition in which it is the minimum distance that two objects can be apart and still be seen as two separate objects. Resolution is normally quoted in Ångstroms (Å $\equiv 10^{-10}$meter).

Another parameter is R_{merge}, an internal measure of the accuracy of the data set. R_{merge} compares the differences between symmetry-related reflections that should be identical in intensity. A usable data set should have an R_{merge} of less than 0.1.

Additional statistical parameters determine the quality of the data set. The ratio of intensity to sigma of intensity (I/σI) gives information about the accuracy of the data set. Percentage of completeness is the number of unique reflections measured over the number of unique reflections possible. If the completeness is greater than 80%, the data will be sufficient for almost all purposes.

4.5 Structure determination

The structure determination of a protein by X-ray diffraction requires the determination of the phase angles for all the reflections generated by the incidence of X-ray on the crystal.

The structure factor F(hkl) for a reflection h,k,l itself is a *complex* number derived as follows :

$$F_{(h,k,l)} = \sum_{j=1}^{atoms} f_{(j)} \exp\left[2\pi \cdot i(hx_{(j)} + ky_{(j)} + lz_{(j)})\right]$$

where:

f_j - is the scattering factor for atom j and depends on the kind of atom and of the atom coordinates (xj, yj, zj).

The electron density of the protein crystal is obtained by calculating the Fourier summation:

$$\rho(x, y, z) = 1/V \sum_{hkl} F(h,k,l) \exp(i\alpha_{hkl}) \exp\left[-2\pi i(hx + ky + lz)\right]$$

where:

α_{hkl} is the phase angle per reflection (h,k,l) and is not available through the diffraction experiment and x, y, z are the coordinates in the unit cell.

To date, there are four important methods to solve the phase problem: molecular replacement (MR), multiple wavelength anomalous dispersion (MAD), multiple isomorphous replacement (MIR) and direct methods for very high resolution diffraction data. Many times a mixture of these methods is used.

4.5.1 Molecular replacement

Molecular replacement was developed by Rossman and Blow (1962) (Rossmann, 1962). In this method the crystallographer use the phases from structure factors of a known protein as initial phases for a new protein. The search model must be correctly positioned with in the unit cell of the structure determined. The search model is positioned in the cell of the unknown structure of rotational and translational transformations. Those transformations can be represented by the next equation (Giacovazzo et al., 1992):

$$X' = [M]X + t$$

where:

X' is the group of vectorial positions of the atoms of the original molecule,

$[M]$ is the matrix that rotates the coordinates X to a new orientation and t is a vectorial translation. Solving the structure by molecular replacement reduces to the problem of finding $[M]$ and t.

To solve the problem we need to use two mathematical functions that are called rotation and translation functions, which do not depend on the knowledge of the unknown structure. Atomic positions in unknown structure are replaced by a map of inter-atomic vectors. Such a map is called Patterson map. The advantage of such a map is that it can be calculate as a Fourier series without phases. The amplitude of

each term is the square of one structure factor, which is proportional to the measured reflections intensity. The Patterson function in the general form is:

$$P(u,v,w)=\frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl}^{2}\cdot e^{-2\pi i(hu+kv+lw)}$$

The coordinates $(u,v.w)$ locate a point in the Patterson map in the same way that coordinates (x,y,z) locate a point in an electron-density map.

The Patterson map is used to initially orient the search model, via the rotation function and then superpose it on the unknown structure, via the translation factor. The application of the rotation matrix $[M]$ and translation vector *t* to the search model leads to an initial model for the structure under investigation and the calculation of a first electron density map.

4.5.2 Multiple Wavelengths Anomalous Dispersion

The multiple wavelength anomalous dispersion (MAD) method was used with success by Hendrickson and his collaborators in the determination of protein structures (Hendrickson et al., 1988); (Raj Lakshmi et al., 1988); (Hendrickson, 1991). The scattering of X-rays due to the valence electrons of atoms differs in phase by exactly 180^0 with the respect to the incident beam. The inner electrons of the atom are more strongly bound to the nucleus, a fact that is especially true for heavy atoms, which have a high nuclear charge. As a consequence, the phase difference between the incident beam and the beam scattered by such electrons is different from 180^0; a property called anomalous scattering. With the MAD technique, all the data can be collected from the same crystal by using several different X-ray wavelengths. Due to the altered scattering properties of the anomalous scattering atom (heavy atom) as a function of wavelength, the intensities of the reflections will change slightly (on the order of 3-5%) from a data set collected on one wavelength to another. The position of the anomalous scatter can be determined from the combined datasets by similar formulas like those used for MIR. In the case of MAD, unlike MIR, the phasing power increase at higher resolutions. The MAD method requires tuneable X-rays, which are available only at synchrotron facilities.

4.5.3 Multiple Isomorphous Replacement

In the multiple isomorphous replacement method (MIR) the phases of the reflections are determined from the knowledge of the position (and therefore, of the phases) of heavy atoms inserted into the proteins crystal. The introduction of the heavy atoms into the crystals results in small differences in intensities for the reflections when compared with those of the crystal without the heavy metal (so called native-crystal).

The position and occupancy of the heavy atoms is the starting point for the determination of the protein phase angles. The difference Patterson function applied for the determination of the heavy atom coordinates through the Patterson peak located at $(u, v.w)$ is:

$$P(u,v,w) = \frac{2}{V}\sum_{h}\sum_{k}\sum_{l} \Delta_{iso}^{2}\cos[2\pi(hu + Kv + lw)]$$

where,

$\Delta_{iso}(hkl) = (F_{PH} - F_{P})$ is the isomorphous difference between the structure amplitude for an *hkl* reflection of protein plus heavy atom F_{PH} and the protein F_P.

4.5.4 Calculation of electron density maps

After the phase determination, an electron density map of the asymmetric unit may be calculated by Fourier transformation of the structure factors. This map is obtained by initially calculating the value for the electron density at points on a three-dimensional lattice. The lattice divides up the asymmetric unit and is usually visualized after countering points on a three-dimensional lattice. The lattice divides up the asymmetric unit and is usually visualized after contouring points of equal electron density to highlight the peaks. These peaks correspond to atomic positions.

The degree of detail to be observed in the structure depends on the quality of the X-ray data and particularly their maximum resolution. Another factor that affects the information present in the map is the degree of order of particular regions of the molecule.

4.5.5 Technique for improvement the electron density maps

The electron density map calculated from the first set of protein phases can be improved through improvement of the protein phase angles. The standard methods used are: omit map, solvent flattening and molecular averaging.

Solvent flattening is the method that introduces the contribution of the solvent molecules in the calculation of the phases. This method assumes that any density in the solvent region of the protein arises from noise fluctuation and that the solvent density should be flat everywhere throughout. The algorithm for the solvent flattening programs developed by Wang (Lindqvist and Branden, 1985) is equivalent to low-pass filtering of data in reciprocal space. The lowest points in this smoothed map are considered to be the solvent. The mask is used to flatten the solvent regions of the map in order to obtain the modified map. This filtered map is inverted by reverse Fourier transform to produce new structure factors. The new structure factors are, then, combined with the original MIR or MAD phase data to obtained filtered structure factors which are used to obtain a new map.

Molecular averaging is the method in which the non-crystallographic symmetry can be used to obtain averaged maps. The electron density maps noise will tend to cancel out and can be used as a phase restriction to improve phasing. The electron density corresponding to the molecules (subunits), related by the non-crystallographic may be essentially identical. The equal density in the molecules related by the non-crystallographic symmetry will improve the structure factors and the phases (Drenth, 1994).

4.5.6 Map fitting

The building of the model and fitting realistically into the molecular surface implied by the map is done using different interactive programs such as “Coot”(Emsley and Cowtan, 2004), XtalView (McRee, 1999), or “O”(Jones et al., 1991). These programs stores automatically the new atoms locations after the model is build or adjusted in a form of three dimensional coordinates. During the process of model building and fitting, it is important to keep in mind that amino acids have a generally fixed and known stereochemestry, also all the bond distances and bond angles are within certain fixed values and with clear boundaries. The degrees of

freedom are only in the rotation about bonds. Use of Ramachandran plots, and data bases of typical bond parameters helps identify poorly built areas of the protein structural model.

4.5.7 Model building

Model building is the process of constructing a three dimensional molecular structure that fit the experimental electron density and, at the same time, maintaining reasonable stereochemistry. Building the model starts with the initial-chain trace of a map. The method in main-chain tracing of the starting model is to use dipeptide like a building block. In general the initial model will contain polyalanine or polyglycine chain. When the best fitting of main-chain to the electron density has been obtained picking and fitting of the side chain can be started. The matching of side chains depends on fitting a pattern of large and small side chain that is unique. If the main chain is fit accurately, then the C-α atom is accurately positioned. At medium resolution the density is a rough guide to the position of the side chain atoms. During the fitting procedure it is necessary to check regularly for the correct stereochemistry.

4.5.8 Model refinement

The initial structural model will contain errors that can be minimized through model refinement. This is a process of adjustment of the atomic coordinates of the model in order to minimize the difference between the experimentally observed structure factors amplitudes (F_o) and those calculated from a hypothetical crystal containing the model (F_c).

Different types of refinement have been introduced, including: rigid-body refinement, simulating annealing, Powell conjugate gradient minimization, refinement of B-factors and atomic occupancy.

For optimization to work well the problem should be highly over-determined, there should be many times as many observations as adjustable parameters. This is generally not the case for protein structure refinement. The crossover point up to which point it would be possible to fit almost anything with the number of parameters is at about 2.8Å. It is possible to make up for the lack of observations by including

more information about the allowed geometry of protein molecules: bond lengths, bond angles, van der Waals contacts, even preferred torsion angles (R. Read, 2006).

Rigid body refinement is the process of refining the position of rigid groups of atoms against the observed amplitudes. All the atoms of the rigid-body will moved in a concerted manner during the refinement, such that the atoms will always occupy the same relative positions. For each rigid-body it is possible to refine up to six parameters, three rotations and three translations. The principle disadvantage of rigid body refinement is the possibility to obtain a local minimum before the correct minimum was obtained. This refinement is the first step for molecular-replacement models.

Non-crystallographic symmetry (NCS) is the symmetry that exists locally in the asymmetric unit of the crystal. This information can be used in order to decrease the number of parameters to be refined. The NCS propose that there is redundancy within the asymmetric unit, for example two or more copies of the same protein subunit. In order to use NCS we assume that all the molecules belonging to the asymmetric unit are structural identical. In the case of two molecules per asymmetric unit the refinement is performed on one molecule plus the parameters that describe the symmetry between the two independent copies.

Simulated annealing was proposed at the end of the 1980's and is a powerful technique for the refinement of the protein structures (Brunger, 1988; Brunger, 1997); (Fujinaga et al., 1989). This technique refers to a process which simulates the dynamics of the protein at high temperature (generally between 2000 and 5000K) followed by gradual cooling. At high temperatures the atoms acquire sufficient kinetic energy to move uphill over energy barriers in the target function. This permits large conformational changes of the structure which provides the possibility for the structure to move from a local energy minimum to a deeper one which is separated by an energy barrier. This technique increases the convergence radius of the refinement, permitting the automatic correction of relatively large errors in the starting model without the need of manual intervention.

The omit map is an electron density map with the observed structure factor amplitude F and with the phase angles α, calculated only for the part of structure that

is known correctly. This method is used to identify the parts of the map which are unclear and can not be traced satisfactorily. The Fourier is possible because all parts of the model contribute to every reflection. The difficult part in the omit map is to use the correct scale factor to scale F_o to F_c. The sum of F_c will be smaller than that of F_o if a portion of the model is omitted. Only about 10% of the model can be left out during calculation of the omit map, so it is necessary to make many omit maps to examine the entire structure.

Unfortunately, the refinement process is not totally automated and must be interrupted periodically to calculate the new electron density maps based on better phases resulting from the new atomic positions. After this these maps are checked by molecular graphics and manual alternations made to the structure. During the refinement process the different types of maps are calculated all the time. In difference maps (F_o-F_c) it is possible to observe positive peaks where the atoms are missing from the model and negative peaks where non-existent atoms have been erroneously included. Many times a pair of peaks, one positive and one negative, is the indication that the particular atom or group of atoms has been slightly incorrectly positioned. Omit maps in which a small region of the structure has been deliberately omitted from the model are also frequently used to determine if the maps and the model are related.

An important parameter that checks the overall quality of the refinement is the discordance index (R-factor) between the observed structure amplitudes (F_o) and those calculated from the model (F_c) (Drenth, 1994).

$$R = \frac{\sum_{hkl} F_o - F_c}{\sum_{hkl} F_o}$$

Another parameter is the free R-factor. The free R-factor is calculated in a fashion similar to that of the R-factor, but only on a partial data set that is not used in the refinement of a structure. Usually 10% of the data are isolated from the working group and after each cycle of the refinement the structure factors are calculated for both groups. The working group is used for the calculation of the R-factor. The test group is used to calculate R_{free} in an exactly analogous manner. The R_{free} is highly

correlated to the phase accuracy of the atomic model and can detect various types of errors. In general, during refinement R_{free} values are higher than R_{work}, but both are supposed to diminish during the refinement, and in the final stages the both often became more similar.

4.6 Functional analysis of the protein targets

The functional analysis of the protein targets before crystallization is an important step. This step is supposed to give us information about the capability of the proteins to preserve their function also *in vitro*. One way to determine the capability of the proteins to function *in vitro* is to check their capability to interact with other proteins. Protein-protein interactions can have a number of different measurable effects. First, they can alter the kinetic properties of proteins. This can be reflected in altered binding of substrates, altered catalysis, or altered allosteric properties of the complex. Few physical methods to detect the protein-protein interaction are:

1. Protein affinity chromatography
2. ITC

1. Protein affinity chromatography. By this method a protein is coupled to a matrix, like Ni or Sepharose, under controlled conditions. The proteins that are supposed to interact with the already immobilized protein pass through the column. If the two proteins interact and connect one to the other the second protein will not be washed out from the column in the washing stage and the bough proteins will leaved the column only in the elution stage.

2. ITC (Isothermal Titration Calorimetry) is a quantitative means available for measuring the thermodynamic properties of a protein-protein interaction. ITC measures the binding equilibrium directly by determining the heat evolved on association of a ligand with its binding partner. In a single experiment, the values of the binding constant (Ka), the stoichiometry (n), and the enthalpy of binding (ΔH_b) are determined. The free energy and entropy of binding are determined from the association constant. The temperature dependence of the ΔH_b parameter, measured by performing the titration at varying temperatures, describes the ΔCp term. An ITC instrument consists of two identical cells composed of a highly efficient thermal conducting material (Hasteloy or gold) surrounded by an adiabatic jacket. In an ITC

experiment, the macromolecule solution is placed in the sample cell. The reference cell contains buffer or water minus the macromolecule. The direct observable measured in an ITC experiment is the time-dependent input of power required to maintain equal temperatures in the sample and reference cell. During the injection of the titrant into the sample cell, heat is taken up or evolved depending on whether the macromolecular association reaction is endothermic or exothermic.

5 Materials and methods

5.1 Materials

5.1.1 General reagents

REAGENT	MANUFACTER
Acetic acid (CH_3COOH)	Bio Lab
Acetone ($(CH_3)_2CO$)	SIGMA
Acrylamide	SIGMA
Acrylamide/bis-Acrylamide 40%	SIGMA
Agar	DIFCO Laboratories
Agarose (electrophoresis grade)	SIGMA
Ammonium Sulphate ($(NH_4)_2SO_4$)	Carlo Erba
Ammonium persulfate ($(NH_4)_2S_2O_8$)	SIGMA
Ampicillin sodium crystalline	SIGMA
Bacto-tryptone	Becton, Dickinson and Co.
Bacto yeast extract	DIFCO Laboratories
Bis Acrylamide (N,N'-Methylene-bis Acrylamide)	SIGMA
Bromphenol blue	SIGMA
Calcium chloride ($CaCl_2$)	Bio-Rad Laboratories
Citric Acid ($C_6H_8O_7$)	FRUTAROM
Coomassie Brilliant Blue R-250	SIGMA
Copper Dinitrate Hexahydrate ($Co(NO_3)2\cdot 6H2O$)	MERCK
Copper Sulphate Pentahydrate ($CuSO_4\cdot 5H_2O$)	MERCK
D,1-Dithiothreitol (DTT)	SIGMA
Dodecyl Maltaside	
EDTA ($C_{10}H_{14}N_2Na_2O_2\cdot 5H_2O$)	Riedel-de Haen
Ethidium bromide (EtBr)	Bio-Rad Laboratories
Ethilen glycol	Hampton Research Corp.
Glycerol (Glycerin)	J.T.Baker
Boric Acid (H_3BO_3)	FRUTAROM
Hepes ($C_8H_{17}N_2O_4SNa$)	SIGMA
Hydrochloric acid 37% (HCl)	Carlo Erba Reagents
Imidazole ($C_3H_4N_2$)	FLUKA
Isopropyl β-D-thiogalactopyranoside (IPTG)	SIGMA,ORNAT
Kanamycine	SIGMA
Lauryl sulphate lithium salt	SIGMA
Lysozyme grade I from chicken egg white	SIGMA

Manganese Chloride ($MnCl_2 \cdot 4H_2O$)	Carlo Erba
Magnesium Chloride ($MgCl_2 \cdot 6H_2O$)	Carlo Erba
Magnesium Sulphate Heptahydrate($MgSO_4 \cdot 7H_2O$)	SIGMA
Mes ($C_6H_{13}NO_4S$)	SIGMA
Methanol	FRUTAROM
Molecular mass standards for SEC-HPLC	SIGMA
Molecular weight markers for SDS-PAGE	SIGMA, Bio-Rad Laboratories
Polyethylene glycol (PEG)	Hampton research Corp.
Potassium dihydrogen phosphate (KH_2PO_4)	MERCK
Q-Sepharose Fast Flow	Pharmacia Biotech
Sigmacote (Chlorinated organopolysiloxane in heptane)	SIGMA
di-Sodium hydrogen phosphate (Na_2HPO_4)	Riedel-de Haën
Sodium Carbonate (Na_2CO_3)	FRUTAROM
Sodium chloride (NaCl)	FRUTAROM
Sodium dodecyl sulfate (SDS)	Biochemical
Sodium hydroxide (NaOH)	FRUTAROM
Sodium Molybdenum Oxide Dihydrate (Na2MoO4·2H2O)	
Sodium Azide (NaN_3)	FLUKA
Sodium nitrate ($NaNO_3$)	MERCK
Sucrose	SIGMA-ALDRICH
N,N,N',N'–tetra-methylethyenediamide (TEMED)	Bio-Rad Laboratories
Tris ($C_4H_{11}NO_3$)	SIGMA
Urea (CH_4N_2O)	SIGMA-ALDRICH
X-gal	SIGMA
Zinc Sulphate Heptahydrate ($ZnSO_4 \cdot 7H_2O$)	

5.1.2 Biochemical, Recombinant DNA Technology and Crystallization Kits

KIT	**MANUFACTER**
CryoPro	Hampton Research Corp.
Crystal Screen™	Hampton Research Corp.
Crystal Screen 2™	Hampton Research Corp.
Index™	Hampton Research Corp.
Plasmid DNA Purification - Midiprep	Nucleobond® PC100 Macherey-Nagel
Plasmid DNA Purification - Miniprep	Nucleospin® Macherey-Nagel

5.1.3 General materials

INSTRUMENT	MANUFACTER
Centricon 30K and 100K	Amicon
Dialysis tubing (molecularporous membrane tubing)	Spectrum, Membrane Filtration products, Inc.
Filters contains nitrocellulose 0.45µm	Millipore Corp.
Gel air cellophane	Bio-Rad Laboratories
Glass Cover Slips	Knittel Glaser
High Vaccuum Grease	Dow Corning
Sterile Needles 0.8mm x 40mm	Becton Dickenson
Sterile Syringe 10ml	Becton Dickenson
Syringe driven filter 0.45µm	Millipore Corp.
Tissue Culture test plates 24 wells	RHENIUM
Tissue-culture 24-well Linbro plate	ICN Inc.

5.1.4 Purification chromatography columns

COLUMN TYPE	COLUMN MATERIAL	MANUFACTURER
Anion Exchange Low Pressure Liquid Chromatography (AEC-LPLC)	DEAE-650S	Toyopearl, Supelco
Anion Exchange High Pressure Liquid Chromatography (AEC-HPLC)	PL-SAX 1000Å	Polymer Laboratories
Size Exclusion High Pressure Liquid Chromatography (SEC-HPLC)	PL-GFC 1000Å	Polymer Laboratories
Nickel Column	Ni-NTA Agarose	QIAGEN

5.1.5 Growth media

Luria Bertani (LB) media: 1% Bacto tryptone, 0.5% Bacto-Yeast extract, 1% MaCl

LB-Agar (for plates): 1% Bacto tryptone, 0.5% Bacto-Yeast extract, 1% MaCl, 2% Agar

Buffer M9x2: 2g/L NH4Cl, 6g/L KH2PO4, 12g/L Na2HPO4

Buffer M9x5: 5g/L NH4Cl, 15g/L KH2PO4, 30g/L Na2HPO4, 2.5g/L NaCl

Mediun component for the production of Se-Met labled protein:

500mg/L MgSO4, 25mg/L FeCl3, 0.5µg/ml Vitamines, 0.16mg/ml Se-Methionine, 80µg/ml, amino acids, 100µg/ml ampicillin, 5mg/ml glucose.

5.1.6 Cell lines

E.coli strain XL1-blue - was used as host for Teasy vector.

E.coli strain M15 – was used as host for the overxpression vector pQE60.

Bacterial starins were grown at 37°C in LB medium and LB agar supplemented with corresponded antibiotics.

5.1.7 Plasmid vectors and oligonucleotides

Vector name	**Cloned gene**	**Stability**	**Inducer**	**Cell line**	**Source**
pGEM®-T Easy	*Tv*-NblA, his/*HspA*, sHSP-syne.	Amp	IPTG, X-gal	DH5α, XL1-blue (for cloning)	Promega
pQE-60	*Tv*-NblA, his/*HspA*, sHSP-syne.	Amp, Kana	IPTG	M15	QIAGENE
pQE-70	*Se*-NblA	Amp, Kana	IPTG	M15	QIAGENE

5.1.8 Vectors description

pQE60 / 70 (QIAGEN): Expression of C-terminal 6xHis tagged proteins in *E.coli* using this vector is based on T5 promoter transcription-translation system.

pGEM®-T Easy (Promega): This vector was used for the cloning of PCR products. The vector contains T7 and SP6 RNA polymerase promoters.

5.1.9 Oligonucleotides

The oligonucleotides were supplied from SIGMA. They were planned such that their sequence contains at least 50% G-C. As follows there are primer sequences.

F_NblA_vulc. : 5’- CCATGGAGCAACGCTTCCCC

R_NblA_vulc. : 5’- AGATCTGGCAGCGCGTTTG

F-HspA_syne: 5’- CCATGGCTCTCATTCTTTAC

R-HspA_syne: 5’- AGATCTGGAAAGCTGAAC

5.1.10 Enzymes:

All enzymes were purchased from New England BioLabs.

5.1.11 Buffers and solutions:

Solutions for electrophoresis of DNA on agarose gel:

TAE running Buffer: 40mM Tris-base, 40mM acetic acid, 1mM EDTA

Agarose gel: 1% agarose in TAE buffer

***Se*-NblA isolation and purification buffers:**

Freanch – press Buffer: 50mM Tris pH=8.

Buffer for extraction of the protein from the pellet: 50mM Tris pH=8 and 2M Urea.

Wash buffer: 50mM Tris pH=8.0, 20mM Imidazole.

Elution buffer: 50mM Tris pH=8.0, 500mM Imidazole.

***Tv*-NblA isolation and purification buffers:**

Freanch – press Buffer: 50mM Tris pH=8.

Wash buffer: 50mM Tris pH=8.0, 20mM Imidazole.

Elution buffer: 50mM Tris pH=8.0, 500mM Imidazole.

Buffer for solubilization of the pure protein: 50mM Tris pH=8 and 2M Urea.

HspA isolation and purification buffers:

Freanch – press Buffer: 50mM Tris pH=8.

Wash buffer: 50mM Tris pH=8.0, 20mM Imidazole.

Elution buffer: 50mM Tris pH=8.0, 500mM Imidazole.

Buffer for solubilization of the pure protein: 50mM Tris pH=8 and 2M Urea.

sHSP-syne isolation and purification buffers:

Freanch – press Buffer: 50mM Tris pH=8.

Wash buffer: 50mM Tris pH=8.0, 20mM Imidazole.

Elution buffer: 50mM Tris pH=8.0, 500mM Imidazole.

SDS-PAGE (according to Laemmli, (Cleveland et al., 1977))

30% Acrylamide: 300g/L Acrylamide, 8g/L bis-Acrylamide

Resolving buffer x 8: 3M Tris (ph=8.8)

Stacking buffer x 4: 0.5M Tris (pH=6.8)

Running buffer x 5: 125mM Tris, 960mM Glycine

Sample buffer x 5: 6% Lauryl sulphate lithium salt, 0.15M DTT,

0.5M Tris (pH=6.8), 30% glycerol, 0.01% bromphenol blue

Staining solution: 0.25% coomassie brilliant blue, 50% methanol, 7% acetic acid

Destaining solution: 10% Methanol, 10% Acetic acid

Gel recipes:

Resolving gel: 14% Acrylamide, 0.4% bis-Acrylamide, 1% APS, 0.1% TEMED in resolving buffer.

Stacking gel: 6% Acrylamide, 0.15% bis-Acrylamide, 2% APS, 0.2% TEMED in stacking buffer

5.1.12 Column material and sources

Columns for analytical size SEC-HPLC: PL-GFC 1000Å, Polymer Laboratories

Bio-Sil® SEC 205-5, Bio-Rad Laboratories

Nickel column: HisTrap™ HP, Amersham Biociences

5.1.13 Crystallization solutions

NblA-vulcanus: 25-30% Ethylene glycol

Se-Met NblA-vulcanus: 25-30% Ethylene glycol; 15% Ethylene glycol, 10% PEG 2000, 50mM MgCl.

NblA-elongatus : 25-30% Ethylene glycol

HspA-vulcanus: 30% PEG 550, 50mM MgCl, 100mM Hepes pH 7.4; 30% PEG 2000, Bis-Tris pH 6.5.

5.2 Methods

5.2.1 Molecular methods

Preparation of competent *E.coli* and transformation

Electroporation method

In this method, *E.coli* species of XL1 Blue and DH5α were prepared as described below.

A starter culture of bacteria was grown overnight at 37°C. 25 ml of the starter was inoculated with 0.5 liter of LB and the culture was grown at 37°C until the culture reached an OD_{600} of ~0.6. The cells were chilled on ice for 15-30 min and pelleted by centrifugation (4000rpm, 10min, and 4°C). The precipitate was resuspended with 1 litter of sterile double distilled cold water and was recovered by another centrifugation (4000rpm, 10min, and 4°C). The precipitate was resuspended with 500 ml of sterile double distilled cold water and was centrifuged again. The precipitate was then gently resuspended with 10% glycerol solution, centrifuged and the precipitate resuspended again to a final volume of 2-3 ml with 10% glycerol solution. The suspension was divided into small aliquots and frozen at -80°C.

Electro-transformation

In the electroporation method, the plasmid DNA is inserted into competent cells by an electric shock that causes modifications in the permeability of cell membrane.

A 50 µl of cell suspension, which was thawed on ice, was mixed with 2µl of DNA. The Gene Pulser apparatus (Bio-Rad) was set at 25µF and 2.5 kV. The Pulse controller was set to 200Ω. The mixture of cells and DNA was transferred to a cold 0.2 cm electroporation cuvette. A single pulse at the above settings was applied. Immediately after the electroporation, the cell suspension was resuspended with 600µl of LB, and was transferred to microcentrifuge tube. As a recovery, the tube was shaken at 37°C for an hour. Following this stage, 100 µl of the cells were plated on plates containing LB, ampicillin and kanamycin.

The Calcium method

In this method, B834 *E.coli* cells were prepared. LB starter of cells which were grown at 37°C overnight was prepared. 30ml of the starter were inoculated with 500 ml LB. the culture was grown at 37°C till OD_{600} ~0.5. The culture was chilled on ice for 10 min and then was centrifuged (6000rpm, 10min, 4°C). The pellet was resuspended with 200 ml of 0.1M $CaCl_2$. After 25 min incubation on ice, the cells were centrifuged (6000rpm, 10min, 4°C). The pellet was resuspended with 13 ml of 0.1M $CaCl_2$ and 7 ml of 60% glycerol. The cells were divided to aliquots of 200µl and were frozen at -80°C.

Transformation

An aliquot of 200 µl of competent cells was thawed on ice. To 100 µl of the cells 2 µl of DNA was added. The mixture was incubated on ice for 30 min and then transferred to 42°C for 90 sec. Following this stage, the cells were chilled on ice for 2 min and 1 ml of LB was added. For recovery, the cells were transferred for 1 hour to 37°C. Then the cells were dispersed on plates containing LB and ampicillin.

Plasmid preparation of *Tv*-NblA and sHSP-synecosystis

Preparation of plasmid DNA (mini and midi prep) was performed according to the manufacture's instructions.

Midi Prep - Nucleobond ® PC100, BioLab

Plasmid isolation of NblA vulc and NblA syne was performed according to original protocol from GFXTMPCR DNA and Gel Band Purification Kit (Amersham Biosciences).

5.2.2 Manipulation of plasmid DNA

DNA digestion by restriction enzymes

Restriction enzymes usage was performed according to manufacture instructions.

In general, plasmid DNA was cut with a 2-5 fold excess of enzyme (units) over DNA (unit/pmol) for 2-3 hours at 37°C.

DNA Ligation

DNA ligations were performed by incubating DNA fragments with appropriately linearized cloning vector in the presence of buffer and T4 DNA ligase. The reagents (DNA, cloning vector, T4 DNA ligase, ligation buffer) were combined in a microcentrifuge tube, and incubated for 3 hours at room temperature or overnight at 4°C.

Agarose gel electrophoresis of DNA

Electrophoresis in agarose gel is used for separating DNA fragments on the basis of their molecular weights. The range of DNA fragments used in this project was resolved on 1% agarose gels. Gels were prepared by boiling the appropriate mass of agarose in 1XTAE, cooling the mixture, adding EtBr to 0.5 μg ml-1, and then setting this in the gel former. All gels were run in 1XTAE. 2.5-20 μl DNA was mixed with 0.5 volume of DNA loading dye and loaded into the gel. Ethidium bromide intercalates DNA, and in this state fluoresces when illuminated by ultraviolet (UV) light. The DNA was visualized by illuminating the gel with UV light on a transilluminator. Gel purification of DNA was performed according to manufactures instructions.

Amplification of DNA by the polymerase chain reaction (PCR)

PCR is an *in vitro* method for enzymatically synthesizing defined sequences of DNA. The reaction uses two oligonucleotide primers that hybridize to opposite strands and flank the target DNA sequence that is to be amplified.
The cycling protocol consists of 25-30 cycles of three temperatures: strand denaturation at 95°C, primer annealing at 55°C and primer extension at 72°C, typically 1 min for each one.

A standard buffer was used for all PCR amplifications. 1 μl of template DNA was mixed with primers at 100 ng/μl. 10mM dNTPs was added to each reaction. The enzyme used was Taq polymerase, (kind gift of Prof. G. Schuster, Department of Biology, Technion).

After the PCR, aliquots of the mixture typically were loaded onto an agarose gel and electrophorisis was performed to detect amplified products.

DNA sequence determination

Sequence determination of DNA was performed by the "Sequence Determination Unit" of the department of Medicine, Technion, using the florescence detection automated DNA sequence analysis method.

5.2.3 Analytical methods

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins separation by electrophoresis was performed according to Laemmli (1970). The principle of this method is based on the fact that protein movement in denaturative gel with SDS, under an electric field, is according to their molecular weights. SDS-PAGE is cast as a resolving gel (14% acrylamide) topped by stacking gel (6% acrylamide). The resolving gel was overlaid with water and some minutes after termination of the polymerization, the water was replaced by the stacking gel mixture. Slab gels were used in a vertical apparatus.

NblA and HspA samples in volume 3-10µl were incubated for 3 min at 80°C in sample buffer, centrifuged for 2 min at 12Kxg, and placed into gels slots. The apparatus was filled up with running buffer containing 0.1% SDS. Gels were run at room temperature at constant current (25mA for each plate). After completion of electrophoresis, the gels were placed in the staining solution and shaken gently. After 15 min the staining solution was poured off and the excess dye was removed by a destain solution, until the gel background was clear. The gels were dried between two layers of cellophane.

5.2.4 Overexpression of the recombinant NblA and HspA proteins.

Overexpression of the His-NblA and HspA proteins

The NblA-elongatus gene cloned in the pQE70 vector which contains a His tag in the end of the gene was a gift from Dr Rakefet Schwartz from Bar-Ilan University. The *Tv*-NblA., *Tv*-HspA and sHSP-syne were cloned in our laboratory in a pQE60 vector contains a His tag in its sequence. As a result we were able to produce the proteins with 6 His residues on their carboxy terminal.

E.coli M15 cells, which contain the plasmid, were grown at 37°C to mid-logarithmic phase in autoclaved Luria-Bertani (LB) broth solution (Sambrook et al. 1989) supplemented with 100mg/ml ampicillin and 50mg/ml kanamycin. A flask, containing

500 ml of LB medium with 100mg/ml ampicillin and 50mg/ml kanamycin, was inoculated with the 10ml overnight culture and allowed to grow until an OD600~0.6 was reached. 1mM of Isopropyl35 β-D-thiogalactopyranoside (IPTG) was then added, and the culture was incubated for further 3-4 hours at 37°C.
Cells were harvested by centrifugation (10000rpm, 10min, 4°C) and were frozen at -80°C until protein purification was performed.

Overxpression of Seleno-L-methionine labeled proteins

In order to produce Se-Met-substituted His-NblA and HspA, the pQE60 and pQE70 was used to transform the methionine auxotroph *E.coli* strain B834, which was provided from the laboratory of Prof. Y. Shoham (Dept. of Biotechnology and Food Engineering,Technion).

A starter culture of the B834 auxotroph strain containing the pQE60 or pQE70 expression plasmid was grown overnight at 37°C in a minimal medium (M9) that contains 4 mg/ml L-methionine.

This starter culture was then centrifuged and resuspended in minimal medium (M9) devoid of methionine but containing 4 mg/ml selenomethionine.
Cells in this medium were grown at 37°C to OD600 of ~ 1.0, at which point IPTG was added as an inducer and the cells were incubated further for 3-4 hours at 37°C. Cells were harvested by centrifugation (10000rpm, 10min, 4°C) and were frozen at -80°C till protein purification was performed.

5.2.5 Isolation, purification and crystallization of the proteins

Isolation and purification of *Se*-NblA for crystallization

Frozen cells (following induction of expression) were resuspended in French press buffer (50mM Tris pH=8) and disrupted by only one passage through the French press. The lysate was then centrifuged (14,000 rpm, 20min, 4°C). The NblA-elongatus was found only in the insoluble fraction and was solubilized by treating the pellet with 2M urea in Tris 50mM pH 8.0. Following treatment with urea, the insoluble fraction was removed by centrifugation (14,000 rpm, 20min, 4°C), leaving partially purified protein in the supernatant. The protein was further purified by metal-chelation chromatography on HisTrap columns (Amersham Biosciences), in an

elution buffer (Tris pH=8, 500mM Imidazole). All the fractions obtained were analyzed by SDS-PAGE. The purest fractions were used for crystallization.

Isolation and purification of *Tv*-NblA, HspA for crystallization

Frozen cells (following induction of expression) were resuspended in French press buffer (50mM Tris pH=8) and disrupted by only one treatment of French press. The lysate was then centrifuged (14,000 rpm, 20min, 4°C). The NblA and HspA proteins were found in the soluble fraction. The proteins were purified by metal-chelation chromatography on HisTrap columns (Amersham Biosciences), in an elution buffer (Tris pH=8, 500mM Imidazole). Following elution, both the NblA and the HspA –vulcanus proteins were insoluble and precipitate immediately or during dialysis against buffer Tris pH 8. The pellet of the insoluble proteins was collected by centrifugation, and was then solubilized in buffer Tris pH 8 with 2M urea. All the fractions obtained were analyzed by SDS-PAGE. The purest fractions were used for crystallization.

Crystallization of NblA proteins

All crystallization experiments were performed using the hanging and sitting drop vapor-diffusion methods in a 24-well tissue-culture plate at 293K. Crystallization drops were formed by mixing 2µl protein with 2µl precipitant, for handing drops, and 7µl protein with 7µl precipitant, for sitting drops, above 1 ml reservoir well solution.

Crystallization of HspA protein

The crystallization experiments of HspA were performed using hanging drop vapour-diffusion method and microbatch under oil (done on the Proteomics Center, Weizmann Institute). For the hanging drop method the crystallization drops were formed by mixing 2µl protein with 2µl precipitant, above 1 ml reservoir well solution. For the microbatch under oil method the crystallization drops were formed by mixing 0.2µl protein with 0.2µl precipitant.

Crystallization screening of NblA and sHSP- synecosystis proteins

For the NblA and sHSP- synecosystis proteins crystallization trials were initially performed using the Crystal Screen I and II kits (Hampton research, USA).

Crystallization screening of HspA proteins

For the HspA the crystallization trials were initially performed in microbatch under oil and hanging drops using the Crystal Screen I and II kits, Index (Hampton research, USA), Wizard I and II (Emerald BioStructures, USA).

Chromatographic analysis

Columns for analytical size exclusion high performance liquid chromatography (SEC-HPLC) were equilibrated with HPLC buffer. The efluants were monitored spectroscopically at 280nm. The column was calibrated using molecular mass standards.

5.2.6 Crystal handling and Diffraction Analysis

Flash freezing technique

Flash frozen crystals are increasingly being used to collect X-ray diffraction data from macromolecules. The technique prolongs the useful crystal lifetime in the X-ray beam so that a complete data set can be collected from a single crystal. Because our crystals grown in relatively high concentration of cryo-protectant (like ethylene glycol, PEG 550, PEG 2000) no incubation with oil or other cryo-protectant was need prior to freezing, and the crystals were directly frozen by plunging them into liquid nitrogen.

Data collection and processing

All X-ray diffraction data collection was performed at the European Synchrotron Radiation Facility (ESRF-Grenoble) beamlines ID14-1, ID14-3, ID-29, BM-30A. Analysis of the diffraction patterns, auto indexing, merging and scaling were performed using HKL (Denzo and Scalepack), Mosflm and XDS program suits.

5.2.7 Computational software

HKL package

The HKL suite (http://www.hkl-xray.com) is a package of programs intended for the analysis of X-ray diffraction data collected from single crystals. It consists of three parts: Xdisplay for visualization of the diffraction pattern in the original form; Denzo for automatic indexing data, Scalepack for merging and scaling of the intensities obtained by Denzo or other programs and global refinement of crystal parameters.

Mosflm

Mosflm (http://www.mrc-lmb.cam.ac.uk/harry/mosflm/) is a program that can integrate single crystal diffraction data from area detectors. It is assumed that the experiment was conducted using the Arndt-Wonacott oscillation method, and also that monochromatic radiation was used. Mosflm can process data from a large variety of modern area detectors.

XDS Program Package

X-ray Detector Software (XDS) (http://www.mpimf-heidelberg.mpg.de/~kabsch/xds/) can process single crystal monochromatic diffraction data. XDS use non-overlapping rotation images collected from a single-crystal at a fixed X-ray wavelength for automatically derived crystal orientation, symmetry, and cell parameters from this data images. XSCALE that puts one or more files obtained from data processing with XDS on a common scale.

BLAST

NblA and HspA proteins were pairwise compared in terms of similar protein sequences in other organisms against the databases using BLAST (Basic Local Alignment Search Tool) sequence comparation NCBI (http://www.ncbi.nlm.nih.gov/BLAST/). BLAST uses an algorithm which seeks local as opposed to global alignments and is therefore able to detect relationships among sequences which share only isolated regions of similarity (Altschul et al, 1990).

ClustalX

Protein sequences were aligned using identity matrix of ClustalX (ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/) with pairwise gap penalties of 10 for gap opening and 0.2 for gap extension.

SHELX

SHELX (http://shelx.uni-ac.gwdg.de/SHELX/index.html) is a set of programs for crystal structure determination from single-crystal diffraction data. SHELX for macromolecular structure determination contains the following programs: SHELXC performs a preliminary setup of files needed for SHELXD and SHELXE. SHELXD performs the actual finding of the location of the selenium structure in SAD or MAD experiments and it is based on dual-space recycling and incorporation of direct methods and Patterson methods (Uson and Sheldrick, 1999). The SHELXD solution with the highest correlation coefficient was input directly into SHELXE (Sheldrick, 2002) for phase calculation and improvement without future refinement of the atom positions or occupancies. The quality of the phase refinement in SHELXE is dependent on the resolution of the native data and the solvent content and generates high-quality maps despite the modest resolution.

Solve/Resolve

SOLVE is a program that can carry out all the steps of macromolecular structure determination from scaling data to calculation of an electron density map, automatically. SOLVE does everything crystallographers do to solve an MIR or MAD structure, but automatically. It scales data, solves Patterson functions, calculates difference Fouriers, looks at a native Fourier to see if there are distinct solvent and protein regions, and can score partial MAD and MIR solutions to build up a complete solution.

RESOLVE is a program that improves electron density maps. RESOLVE uses a statistical approach to combine experimental X-ray diffraction information with knowledge about the expected characteristics of an electron density map of a macromolecule.

CNS and CCP4

Crystallography & NMR System (CNS) (http://cns.csb.yale.edu/v1.1/) and CCP4 (http://www.ccp4.ac.uk/) are the result of an international collaborative effort among several research groups. The program has been designed to provide a flexible multi-level hierarchical approach for the most commonly used algorithms in macromolecular structure determination. Highlights include, experimental phasing (including MAD and MIR), molecular replacement, density modification and crystallographic refinement with maximum likelihood targets.

PHASER

Phaser is a program for phasing macromolecular crystal structures with maximum likelihood methods. It has been developed by Randy Read's group at the University of Cambridge and will be available through the Phenix and CCP4 software suites. Intuitively, the principle of maximum likelihood means that the best model is the one that agrees best with the data. Agreement is measured by the probability of measuring the data. So likelihood is the probability that the measure set of observed data, given whatever model you have of the experiment.

XtalView

XtalView (http://www.sdsc.edu/CCMS/Packages/XTALVIEW/xtalview.html) is a collection of programs for solving a macromoleclar crystal structure by different methods; including isomorphous replacemen and MAD, including building of the molecular model, map fitting and manual refinement possibilities. Xfit is a model-building and map viewing program in XtalView (McRee, 1999) which is used for map calculations including the creation of omit maps and the updating of structure factors to reflect model changes from within the program.

COOT

Coot (http://www.ysbl.york.ac.uk/~emsley/coot/) is a programme for model building, model completion and validation. Coot can displays maps and models and allows model manipulations such as idealization, real space refinement, manual

rotation/translation, rigid-body fitting, ligand search, solvation, mutations, rotamers, Ramachandran plots.

Molecular Graphics Programs

The output structure files were displayed with molecular graphics programs such as RasMol (http://www.umass.ed u/microbio/rasmol/), Swiss-pdb Viewer (http://www.expasy.org/spdbv/) and PyMol (http://pymol.sourceforge.net/). These programs also perform a visual analysis of proteins in order to deduce structural alignments and compare their active sites or any other relevant parts. The electrostatic potential of the protein surfaces were calculated using the algorithim imbedded in Pymol, using vacuum conditions. The semi-quantitative nature of this calculation is valid for the comparison of overall potentials between similar proteins.

6 Results and discussions

6.1 Cloning of the *NblA* and *HspA* genes

The pQE60 vector system provides a method for expressing and purifying a protein produced from a cloned gene with a 6xHis tagged on its C terminus. We chose to clone the *Tv-NblA* and *HspA* genes from *T. vulcanus* genomic DNA. Following purification of *T. vulcanus* DNA, we used the polymerase chain reaction (PCR) for the amplification of the *Tv-NblA* and *HspA* genes. We used a forward oligonucleotide (for details see Materials and Methods) incorporating a NcoI site on the 5'end of the gene, and two reverse oligonucleotides (for details see Materials and Methods) incorporating a BglII site at the 3'. PCR was performed using Taq DNA polymerase enzyme, which synthesizes DNA *in vitro* with reasonable fidelity. The taq DNA polymerase amplification products were digested simultaneously with the restriction enzymes NcoI and BglII. We purified the DNA fragments including the *Tv-NblA* and *HspA* genes by using the GFXTMPCR DNA and gel Band Purification Kit. The insert was cloned into the pGEM-T-Easy cloning vector, which can easily ligate all PCR products.

There were two reasons for performing for this step. The first one is that it is a convenient and safe way to excise the insert. The second reason was to be able to obtain the entire sequence of the gene, to ensure the fidelity of the gene duplication by PCR. Selection was performed by the analysis of blue-white colonies. White colonies were picked, DNA was isolated and the plasmids were digested with BglII and NcoI and analyzed by DNA agarose gel electrophoresis. The *Tv-NblA* and *HspA* gene inserts were then cloned into the pQE60 expression vector. The *Se-NblA* gene was a gift from Dr. Rakefet Schwarz (Bar-Ilan University) and was cloned in the pQE70 expression vector. By transformation by electroporation, we inserted all of the plasmids containing the *NblA* and *HspA* genes into M15 *E. coli* host cells. The selection for colonies which contained the plasmid was performed by ampicillin and kanamycin.

6.2 NblA and HspA isolation and purification

Generally, solving the structure of a protein can help to elucidate its function(s) and allows the rational design of future experiments aimed at elucidating the entire chemical characteristics of the molecule. We have used X-ray crystallographic analysis in order to determine the structures of *Se*-NblA, *Tv*-NblA, *Tv*-HspA. Protein structure determination involves following steps which must be performed sequentially:

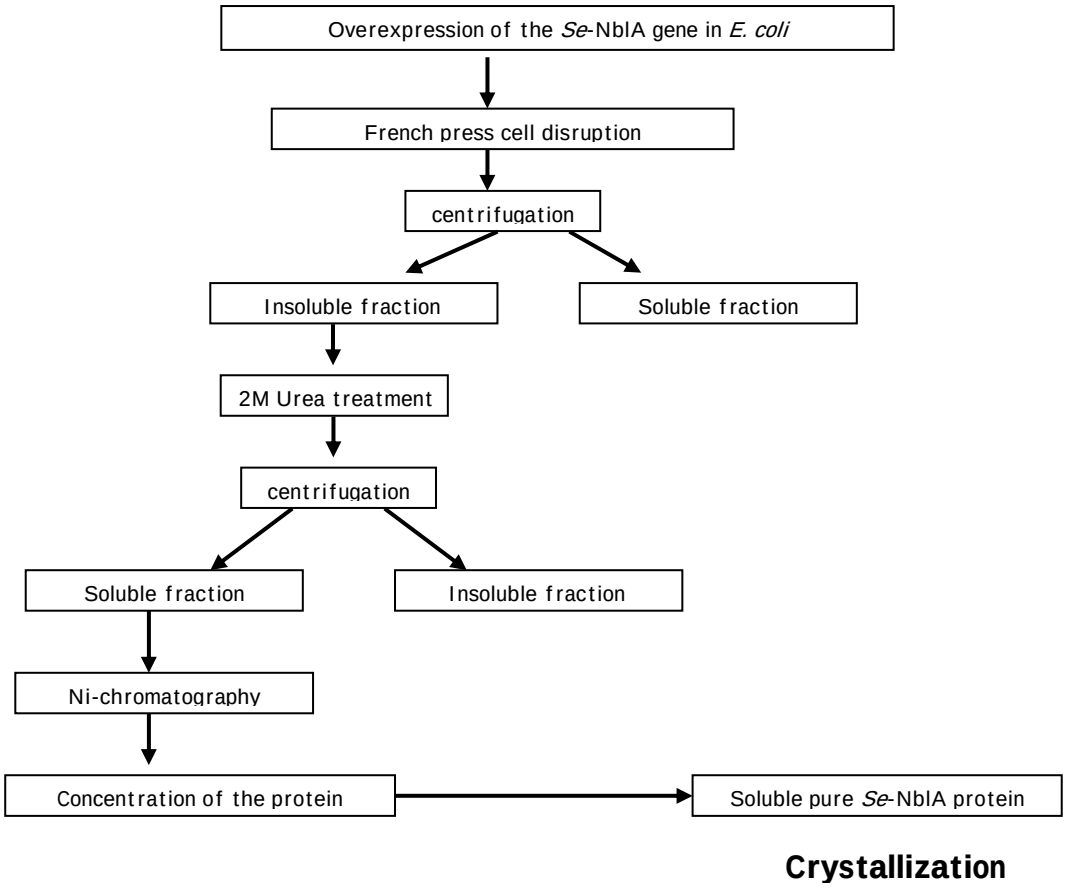
- Preparation of highly purified and soluble protein samples
- Crystallization
- Measurement of diffraction data
- Solving the phase determination problem using molecular replacement or natural or artificially incorporated heavy atoms
- Map interpretation and model building
- Model refinement

Overexpression of protein is a method used to produce a large quantity of a particular protein for biochemical studies or for commercial purposes by increasing the number of molecules of the desired protein per cell. To obtain the recombinant NblA proteins we used genes identified and cloned from *Synechococcus elongatus* sp. PCC 7942 and from *T. vulcanus* and for the recombinant HspA we used the gene identified and cloned from *T. vulcanus.*

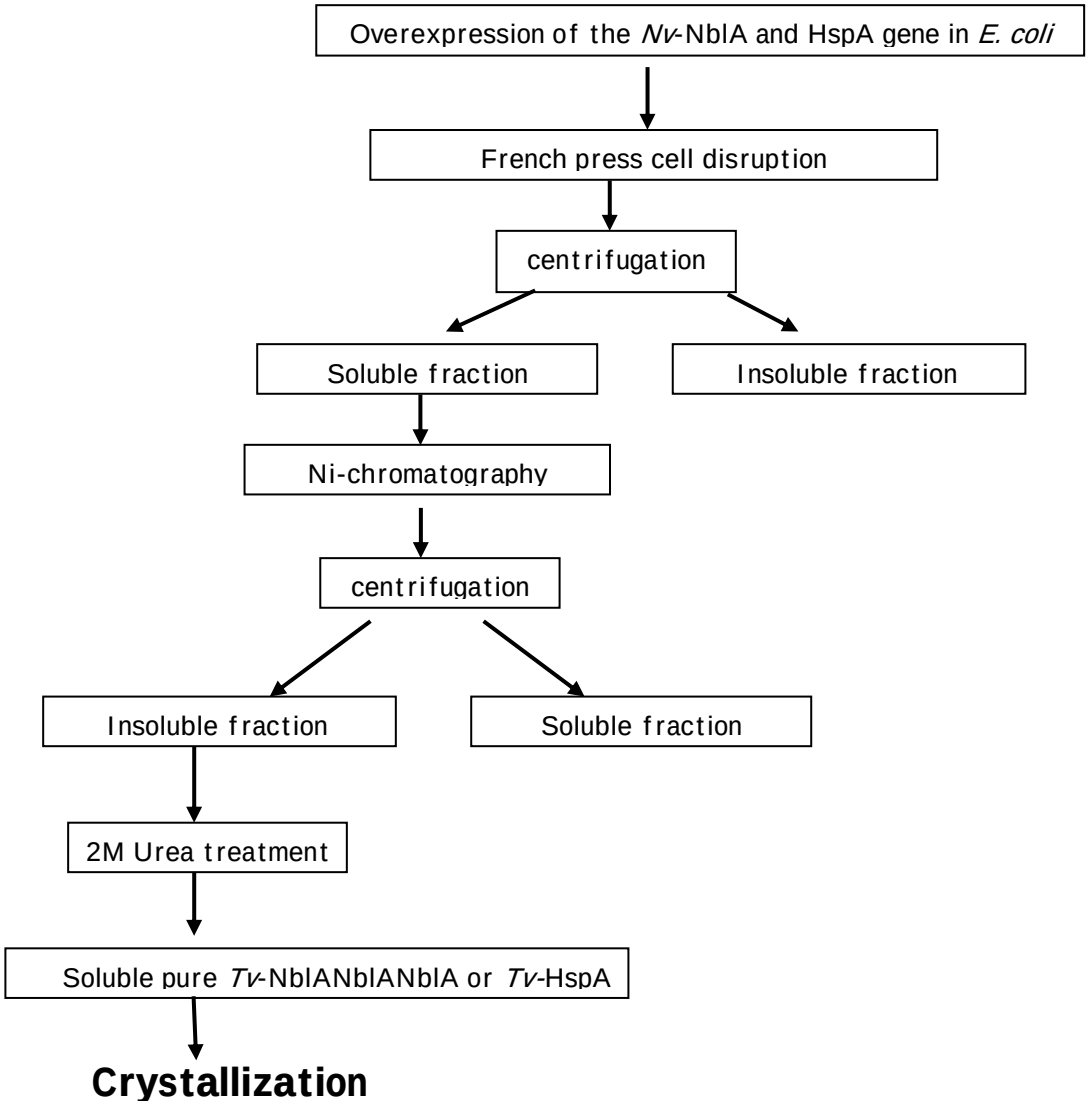
Genetic manipulation allows the addition of affinity tags to expressed proteins, such as a His-tag or a fusion protein such as glutathione S-transferase which is used for efficient purification of the recombinant protein. A His-tag includes the attachment of six or eight histidine residues at the N- or C- terminal of a protein and allows the purification of the protein using a nickel-chelating resin and a glutathione S-transferase subunit at the N- or C- terminal allows the protein to bind to gluthatione columns. The *NblA* and *HspA* genes were cloned to include a His-tag at the C-terminal of the proteins.

Crystallographic analysis requires large amounts of preferably highly purified soluble material. The original amount of soluble protein which is used for each crystallization trial is usually around 1.5-2 mg at 97-99% purity. The basic steps of

preparation of soluble NblA and HspA proteins are show in scheme 1 and 2. Use of those protocols enabled the isolation and purification of significant amounts of soluble NblA and HspA suitable for crystallization from 250ml growth media.



Scheme 1. Preparation of Se-NblA soluble protein for crystallization.



Scheme 2. Preparation of soluble *Tv*-NblA and *Tv*-HspA for crystallization.

The NblA and HspA proteins are expressed at a high level in the *E. coli* host. (see figure 11)

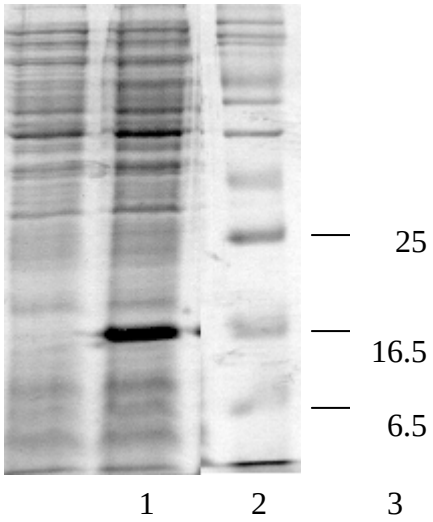


Figure 11. The expression of *Tv*-HspA in M15[pREP4] *E.coli* host. Lane 1, total noninduced crude cell lysate. Lane 2, total induced crude cell lysate 0.5 hour after initiation of expression. Lane 3 molecular mass markers (kDa).

SDS-PAGE was used to analyze protein composition, purity and molecular weight of the NblA and HspA during overexpression, isolation and purification. This technique gives information of the degree of chemical purity and an estimate of the molecular weight of the protein subunits, although nothing can be concluded regarding the three-dimensional integrity of the molecule.

SEC-HPLC was tried to be used to get information about molecular weight of NblA and HspA, oligomeric state and there homogeneity. In order to determine proteins secondary structures we tried to use UV spectroscopy (including the use of circular dicroism for the quantification of secondary structure in the 190-240nm region) but this was difficult because of the presence of urea (Kelly et al., 2005). These techniques were used following the final stage of purification.

Cellular disruption is the first step in protein isolation and one of the most critical steps affecting yield and quality of the isolated protein. Typically, cell disruption needs to be fast and thorough. We used only mechanical disruption method. Mechanical methods are based on the fact that cell envelope is physically broken, releasing all intracellular components in the surrounding medium. For this we used the French Press method, where the cells are physically torn by extrusion through a small aperture at high mechanical pressure.

Different buffers (Tris, Mes, Hepes) were tested from the beginning in order to immediately obtain large amounts of soluble recombinant proteins after disruption. We also tried the addition of different chelating, DTT, EDTA. The different buffers and chelating agents did not make any difference in the disruption process or in the yield of soluble proteins and we decided to continue with 50mM Tris buffer pH=8.0. We tried also the addition of lyzozyme, which cleaves a bond in the peptidoglycan layer of the *E. coli* cell wall (thereby facilitating cell disruption), but this did not influence the yield, and so we did not continue to use this treatment.

All the recombinant proteins obtained have some solubility problems. Therefore we have developed a simple protocol whereby the presence of a significant concentration of urea enabled us to concentrate the target proteins without denaturation.

Overexpressed recombinant *Se*-NblA protein was found only in the insoluble fraction after cell disruption and purification. The protein could be solubilized by treating the pellet with 2M urea in Tris 50 mM pH 8.0. It was thus probably not in an inclusion body, rather it associated with the cell debris and membranes. It is also

possible that the *Se*-NblA protein have minimal incorrectly folded regions. Following treatment with urea, the insoluble fraction was removed by centrifugation, leaving partially purified protein in the supernatant. The protein could be purified further by metal-chelation chromatography as above. In most cases the protein remained soluble in the absence of urea.

Tv-NblA and *Tv*-HspA protein were found in the soluble fraction following cell disruption and centrifugation. All proteins were then purified by metal-chelation chromatography on HisTrap columns (Amersham Biosciences), in an elution buffer containing 500mM imidazole in 50mM Tris buffer pH=8.0. However, following elution, both the *Tv*-NblA and the *Tv*-HspA were insoluble and precipitate immediately or during dialysis against buffer Tris pH=8.0. The pellet of the insoluble proteins was collected by centrifugation, and was then solubilized in buffer Tris pH=8 with 2M urea.

The results of the metal-chelation chromatography was analyzed by SDS-PAGE and can be observed in figure 12.

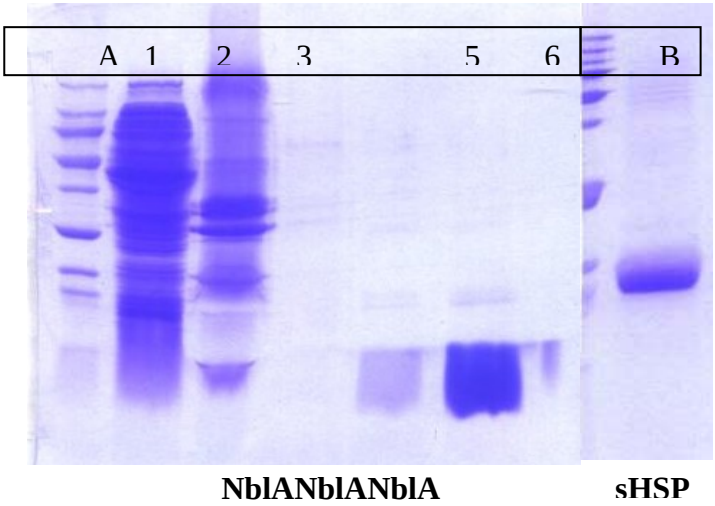


Figure 12. SDS-PAGE of the purified proteins prior to crystallization. A. Purified NblANblANblA. Line 1 – proreins markers. Line 2 - the proteins extract before purification. Line3 – the flow through. Line 5,6 – the pure NblANblANblA protein. B. Purified sHSP

6.3 NblA and HspA solubility problems

6.3.1 Methods to solve the solubility problems in HspA

The first method that we tried was aimed at to produce different kinds of chaperons into the *E. coli* cells by osmotic stress. These chaperons and a very slow induction time were designed to help the HspA to fold in a better way that will confer an increase in protein solubility. For these reasons, *E. coli* cells harbouring the PQE-60/HspA-His plasmid were grown at 37°C in LB medium containing 50 μg/ml of ampicillin, 1M sorbitol, 2.5 mM Betain. At an OD_{600}=0.6, the HspA protein was induced with 0.1 mM of IPTG for 40-60h at 18°C. However, this method did not improve the HspA solubility.

A second method was to denature the HspA after elution with 8M urea followed by rapid dialyses of the solution against a large volume of 50 mM Tris-HCl pH 8.0. This method was used to try to unfold and refold the misfolded variants of the expressed protein. However, the protein precipitated during the dialysis.

A third method was to add different reagents, such as detergents, EDTA, DTT, various salts; however none of these treatments were able to positively influence the solubility of the protein. We tried to change the temperature used during the purification process and we also used buffers with different pHs (varied from 4-10). However, none of this was able to improve the protein solubility.

6.3.2 Addition of 2M urea solves the NblA and HspA solubility problem

As we described previously the NblA and the HspA recombinant proteins have solubility problems when isolated from *E. coli* host cells. Lack of solubility is typically a measure of the formation of non-specific interactions between protein molecules as opposed to those between protein and solvent. In the literature are reports of various measures that have been proposed to alleviate this problem, including the addition of non-ionic detergents and low concentrations of chaotropic agents. It is clear however that one must avoid application of treatment that would result in protein denaturation. Thus the concentration of solubilizing additives used is

typically low. On the other hand, crystallization requires the specific association of the protein molecules by either ionic, polar, hydrogen-bond or hydrophobic interactions. Addition of solubilizing agents should prevent non-specific aggregation without interfering with the formation of the specific contacts needed for lattice formation.

As we described previously in the case of HspA we tried most of the methods already described in the literature, however none the treatments were able to positively influence the solubility of the protein.We decided to attempt to find the maximum concentration of the chaotropic reagent urea that will provide the surrounding required for protein solubility, yet not prevent the association required for crystallization. For this we used buffers containing 2M urea. The mode of interaction between urea and soluble proteins has been extensively studied (Bolen, 2004; Bolen and Baskakov, 2001; Street et al., 2006). The denaturing effect of urea has been proposed to primarily be a result of the association of urea molecules with the backbone of proteins, thereby forcing typically buried residues to interact with the solvent. At relatively lower concentrations, urea can apparently serve as a solubilizing agent, which can even lead to crystal formation (Bolen, 2004). Minimization of the urea concentration from 2M to 1M did not give any significant changes in the crystals quality.

In the end, we have been able to obtain diffraction quality crystals of proteins in this manner.

6.4 Crystallization of NblA

Crystallization trials using the maximal concentration of soluble *Tv*-NblA protein that could be obtained without the presence of urea did not result in the growth of crystals, in most cases. We thus initiated the screening of possible crystallization conditions in the presence of urea, at protein concentrations of about 1mg/ml. For the NblA proteins crystallization trials were initially performed using the Crystal Screen I and II kits (Hampton research, USA). NblA protein crystallization was performed using the vapor-diffusion method. 2 μl of protein (1 mg/ml for the both proteins) was mixed with an equal volume of crystallization reservoir solutions. Crystals of *Tv* and *Se*-NblA protein were obtained in the presence of urea and also in

its absence with 25% ethylene glycol as the precipitating reagent (Figure 13). In the case of *Tv*-NblA we were also able to obtain Se-Met derivates crystals with urea. Crystals that were grown using sitting and hanging drop methods were stabile in the same way and they continue to retain their diffraction quality for at least 2 years. The crystals obtained were very small; the maximum size was about 0.1X0.04X0.04 mm (Figure 13).

In the case of urea solubilized *Tv*-sHSP the crystallization trials were initially performed in microbatch under oil and handing drops using the Crystal Screen I and II kits, Index (Hampton research, USA), Wizard I and II (Emerald BioStructures, USA). Crystals were obtained in more than 20 different conditions, most of them including various PEGs. The best crystals grown in hanging drops in 30% PEG MME 550 (Figure 13) and 30% PEG 2000. The crystals obtained were also very small; the maximum size was about 0.1X0.04X0.04 mm (Figure 13).

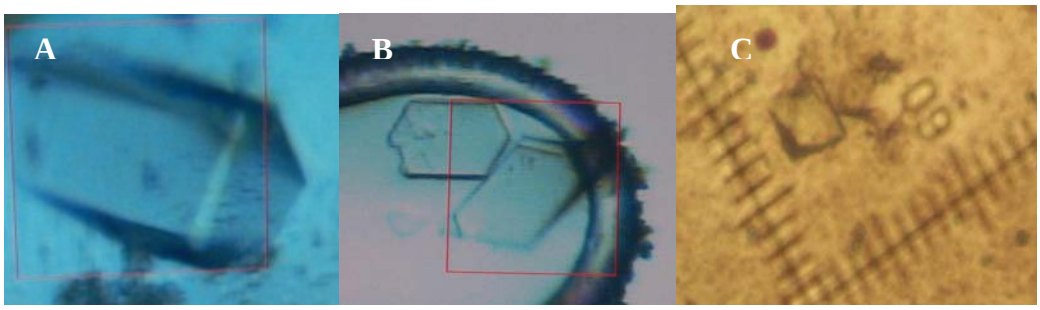


Figure 13. A. Crystal of NblA from *S. elongatus* sp. PCC 7984 grown in the presence of 2M urea and ethylene glycol. The crystals dimensions are about 0.1x0.05x0.04mm. B. Crystals of NblA from *T. vulcanus* grown in presence of 2M urea and ethylene glycol, mounted in nylon loop. The crystal dimensions are about 0.06x0.04x0.01mm. C. *Tv*-sHSP crystal grown in the presence of 2M urea and PEG MME 550. The crystals dimensions are about 0.04x0.04x0.04mm.

Many attempts at optimization of the crystallization conditions which would increase the crystals size were unsuccessful. We tried to add different additives (more than 50) to the crystallization drops that were supposed to influence the crystallization of the protein. Another possible way to improve and increase the size of the crystal is seeding. Seeding is a method of insuring nucleation and crystal growth. Commonly a protein crystal or crystalline aggregate is placed in a storage solution and then crushed

to a fine powder (producing a large number of microscopic crystals) and mixed to make a seed solution. The seed solution is then added to the supersaturated crystallization trials (e.g. added to the equilibrated hanging and sitting drop) thus ensuring the presence of nuclei for crystal growth. The last way that we tried is “backing off” from normal nucleation conditions (Saridakis, E. and Chayen, N.E., 2000, 2003). In this method the drops were incubated for some time at conditions normally giving many small crystals and after that the reservoir was replaced to the reservoirs at concentrations that normally yield clear drops. This can result in much larger and better diffracting crystals as, compared with conventional crystallization at similar conditions. Using either of these methods no increase in the size of the NblA and sHSP crystals was obtained.

6.5 Preliminary X-ray Diffraction Analysis and Data Set Collection

Preliminary information about crystal size and their quality was obtained by analysis under a polarizing microscope. This visual test determines the degree of suitability of the crystals for data collection. Relatively large, well-formed crystals were mounted and examined for diffraction quality. A quick step is made to determine diffraction pattern and resolution limits. A data set was collected only from crystals which passed the preliminary checks.

All the crystals were too small and fragile to be diffracted at room temperature with our in-house diffractometer and were thus analysed by synchrotron radiation. All the data sets were collected using the method of flash freezing. This method uses a cold nitrogen stream at a temperature of around 100 K. Protein crystals are composed of large fractions of liquid with wide channels (~ 10-100 Å) filled with the crystallizing solution. When using low temperatures for X-ray data collection, the solution in the protein crystal channels and mother liquor surrounding the crystal must be vitrified (transformed into an amorphous state). With crystalline ice formation, the solution within protein channels would expand, destroying the protein crystal lattice. Therefore protein crystals are ‘flash cooled', rapidly brought to cryogenic temperatures, in a cryogenic stream or by plunging in liquid nitrogen before data collection. The crystallization conditions included ‘cryoprotectants' such as PEG and

ethylene glycol in the mother liquor and the crystals and so probably reduced cooling rates required to achieve vitrification. In general, data quality of liquid plunged crystals was better than that of gas cooled crystals. The frozen crystals were mounted on the goniometer and in a cold nitrogen stream. The sample was optically centered with the aid of a video camera such that no translations were observed as the crystal was rotated to all angles.

X-ray diffraction data from native *Tv*-NblA and *Se*-NblA crystals were collected at ESRF using ID-14-1 with a monochromatic X-ray beam of fixed wavelength 0.934Å (13.270keV), a beam size of 50 to 200 microns, ADSC Q210 CCD detector and an Oxford Cryostream for 100K data collection. X-ray diffraction data from native HspA crystals were collected also at ESRF using ID-29 with an ADSC Q210 2D detector which has an active area of 4096*4096 pixels, which are 51.2*51.2 microns in size. The detailed crystal parameters and data collection statistics are shown in Table 1A for *Tv*-NblA, Table 1B for *Tv*-NblA without urea, Table 1C for *Se*-NblA and Table 1D for HspA.

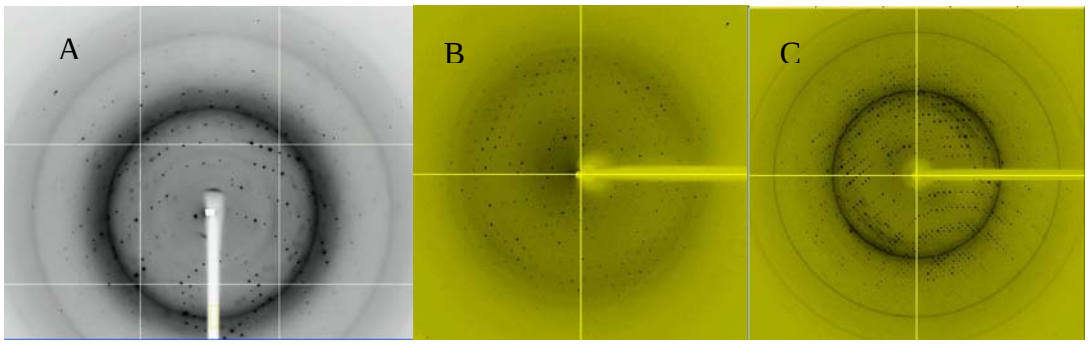


Figure 14 a) 1° oscillation diffraction pattern to 2.5Å collected on beamline ID-29 at the ESRF, Grenoble. b) 1° oscillation diffraction pattern to 2.5Å of *Tv*-NblA collected on beamline ID-14-1 at the ESRF, Grenoble. c) 1° oscillation diffraction pattern to 2.2Å of *Se*-NblA collected on beamline ID-14-1 at the ESRF, Grenoble.

Table 1. Native diffraction data sets statistics. A) Summary of *Tv*-NblA with crystal parameters and data-collection statistics. B) Summary of *Se*-NblA crystal parameters and data-collection. C) Summary of HspA crystal parameters and data-collection.

A)

Crystal	*Tv*-NblA with urea	*Tv*-NblA without urea
Beam line	ID14-1 (ESRF)	ID14-1 (ESRF)
Wave length (λ)	0.9792	0.9792
Space Group	$P3_221$	P1
Unit Cell dimensions (Å,^{O})	a=b=43; c=149;	a=42.19; b=42.39; c=50.56;
	α=β=90; γ=120.	α=69; β=83.36; γ=61.25.
Resolution Range (Å)	49-2.5	47-2.5
No. of Reflections	37675	9900
No. of Unique Reflections	6108	4599
R_{merge}	0.073 (0.36)	0.09 (0.11)
Completeness (%)	99.8 (99.8)	96.7 (93.2)
Multiplicity	6.2 (5.9)	2
I/σ(I)	19.1 (2.3)	19.48 (3.47)

B)

Crystal	*Se*-NblA without urea
Beam line	ID14-1 (ESRF)
Wave length (λ)	0.9792
Space Group	P4
Unit Cell dimensions (Å,^{O})	a=b=78; c=70; α=β=γ=90.
Resolution Range (Å)	25-2.2
No. of Reflections	104112
No. of Unique Reflections	15317
R_{merge}	0.057 (0.35)
Completeness (%)	99.8 (100)
Multiplicity	6.8 (3.5)
I/σ(I)	9.75 (2.1)

C)

Crystal	HspA with urea
Space Group	P3
Beam line	ID 29 (ESRF)
Wave length (λ)	0.9765
Unit Cell dimensions (Å,^{O})	a=b=45.3; c=53; α=β=90; γ=120.
Resolution Range (Å)	53-1.9
No. of Reflections	46314
No. of Unique Reflections	11113
R_{merge}	0.06 (0.26)

Completeness (%)	98.2 (100)
Multiplicity	4.2 (8.7)
I/σ(I)	14 (2)

Rmerge = ((ΣhΣi|I(h)i -<I(h)>)/(ΣhΣi|I(h)i))

6.6 Data set Analysis

Integrated intensity information for each reflection was obtained by reducing of the data frames with three programs DENZO (Otwinowski, Z. and Minor, W., 1997), XDS (Kabsch, 1988, 1993), Mosflm (A. G. W. Leslie, 2006). Finally the data was merged and scaled to produce a suitable data set by SCALEPACK and SCALA.

Autoindexing and scaling of the data and analysis of the pattern of systematic absences give information about the unit cell. A unit cell is characterized by two values, namely by the crystal system and by the space group. These specify how the molecules within a unit cell are related to one another and how the unit cell itself is related to the structure of the whole crystal.

Preliminary analysis using DENZO indicated that the *Tv*-NblA crystals belonged to the $P3_1$ space group with a partial merohedral twinning factor of 0.48. However a further analysis of the data convinced us that the true space group is $P3_221$ and that the data was untwinned. The factors taken into account were the merging statistics and the success of molecular replacement. Analysis of the diffraction pattern of the *Tv*- NblA without urea indicates that the crystals belong to P1 space group.

In the case of *Se*-NblA crystals the most probable space group is P4 with a partial merohedral twinning factor of 0.48 indicated also by Mosflm and by Denzo as well as the Yeates server (http://nihserver.mbi.ucla.edu/Twinning/) The Partial Merohedral Twinning Test developed by Yeates utilizes a statistical comparison of twin-law-related reflection pairs. Without loss of validity, the specified percentage of these pairs are discarded based upon the sum of their intensities in order to minimize statistical error (Yeates, 1997). For the *Tv*-HspA the space group was determined by Mosflm to be P3.

The first step in macromolecular structure determination is estimating the number of the molecules in the crystallographic asymmetric unit. Matthews (Matthews, 1968, 1976) defined V_M (Matthews coefficient) as the density of the protein structure in cubic angstroms per Dalton. Normal values are between 1.8 (tightly packed, little room for solvent) and 3.5 (loosely packed, much space for solvent). Consideration of the possible values of V_M suggest that the asymmetric unit of *Tv*- NblA contain a dimer with a V_M of 2.9 $Å^3$/Da, *Tv*- NblA without urea contain a tetramer with a V_M of 2.64 $Å^3$/Da, *Se*-NblA contain a hexamer or a octamer with a V_M of 2.56 $Å^3$/Da or 1.92 $Å^3$/Da, *Tv*-HspA contain only a monomer with a V_M of 1.9 $Å^3$/Da.

6.7 Structure Determination of *Tv*-NblA

6.7.1 Selenium Multiwavelength Anomalous Dispersion of Se-Met-*Tv*-NblA

Selenium and sulphur are in the same family in the periodic table, which allows for selective replacement of sulphur with selenium. The anomalous signals may be found as location of the K-edges for the selenium. The curves on Figure 15 represent the theoretical values of the dispersive anomalous coefficients f' and f'' as a function of X-ray energy (1Å=12398eV). The theoretical value for the Se K-edge is 12657.8eV (0.9795 Å).

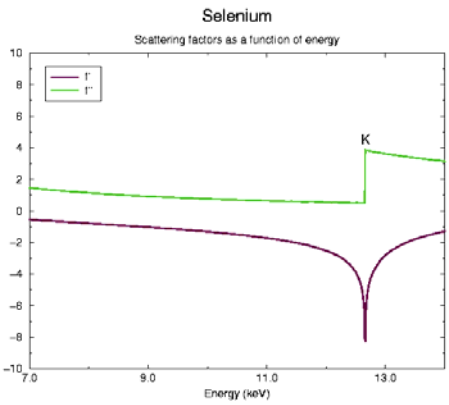


Figure 15. Theoretical location of K-edges for selenium.

MAD experiments were conducted to evaluate the occupancy of selenium at the Se-methionine labelling *Tv*-NblA. Before data collection, a fluorescence scan centred around the Se K-edge was recorded in order to accurately tune the X-ray beam to the

three wavelengths (figure 16), i.e. the absorption maximum (peak) with λ=0.97913Å, the inflection point (edge) with λ=0.97570Å, and a last wavelength of 389eV above the peak wavelength (remote) with λ=0.97970Å. All three MAD crystallographic data sets were collected from single frozen crystals using synchrotron radiation at the beamline BM-30 and ID14-4 in ESRF. The three wavelengths were alternated after each sweep of 180^0 in phi with 1^0 oscillation. All data were integrated and merged using DENZO and SCALEPACK or using Mosflm and SCALA, respectively. Selected statistics of the anomalous data collection are listed in Table 2.

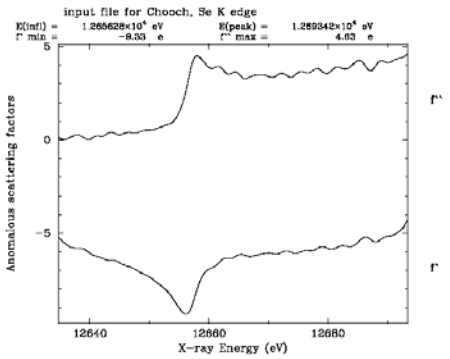


Figure 16. X-ray fluorescence spectrum of the Se-Met-*Tv*-NblA crystal around the Se K- edge (1Å=12398.5 eV).
Data collected on beamline BM-30 in ESRF.

Table 2. Summary of Se-Met-*Tv*-NblA crystal parameters and data collection statistics.

Crystal	*Tv*-NblA Se-Met with urea		
Beam line	BM-30 (ESRF)		
Space Group	$P3_221$		
Unit Cell dimensions	a=b=43; c=149; α=β=90; γ=120.		
	λpeak	λinflection	λremote
Wave length (Å)	0.97913	0.97570	0.97970
Resolution Range (Å)	49-2.5	49-2.5	49-2.8
No. of Reflections	69468	36709	40265
No. of Unique Reflections	6028	4048	4541
Rmerge	0.075 (0.267)	0.08 (0.7)	0.084 (0.385)

Completeness (%)	99.9 (100)	99.9 (79.6)	99.9 (100)
Multiplicity	11.5 (11)	9.1 (2.8)	8.9 (9.2)
I/σ(I)	24.2 (1.3)	21.5 (0.6)	19.7 (1.9)

The selenium structure and initial phases was tried to be determine using SHELEXD/E (Sheldrick and Schneider, 1997) and Solve/Resolve (Terwilliger, T.C. and J. Berendzen., 1999; Terwilliger, T. C., 2000; Terwilliger, T.C., 2002). Unfortunately, the initial electron density map obtained did not able the build of the protein structure. Therefore we tried to determine the *Tv*-NblA structure by molecular replacement.

6.7.2 Molecular Replacement

Another way to solve he phase problem is to possess an atomic model, from which estimate of the phases can be computed. Molecular replacement is the technique that uses of a known structure of a related protein. Most of the times the level of resemblance of two protein structures correlates well with the level of sequence identity.

The molecular structure of NblA from Anabaena sp. PCC 7120 (Bienert et al., 2006) (PDB code 1OJH) was used as a search model for molecular replacement. In the Anabaena NblA crystals are 12 molecules in the asymmetric unit. For the molecular replacement we used only one monomer. The level of homology on the amino acid sequence level between the Anabaena NblA and *Tv*-NblA is about 20% identity and 40% similarity. Most of the homologous residues are spread out over most of the protein sequence, and not clustered in certain regions. The final structure used as the additional search model was between the residues 7-54 and was formed only by two α-helixes coupled by a short loop (see figure 17).

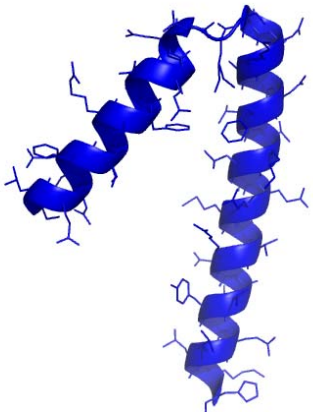


Figure 17. Homology based molecular model of NblA protein folded according to structure of Anabaena NblA

Phasing by molecular replacement was performed using the program Phaser (Read, 2004) from the CCP4 suit of programs (Collaborative Computational Project, Number 4, 1994). This program allows the searching after multiple molecules in the asymmetric unit by maximum likelihood methods. The molecular replacement gives the best LOG-LIKELIHOOD GAIN of 110 score using a data subset between 12 Å and 2.75Å resolution. The RMS, coordinate error expected for the model, used for the molecular replacement was 1.2.

Using Phaser we were able to obtain an initial *Tv*-NblA model formed from a dimer similar to the Anabena dimers (see figure 18) that convinced us that the molecular replacement solution is indeed correct.

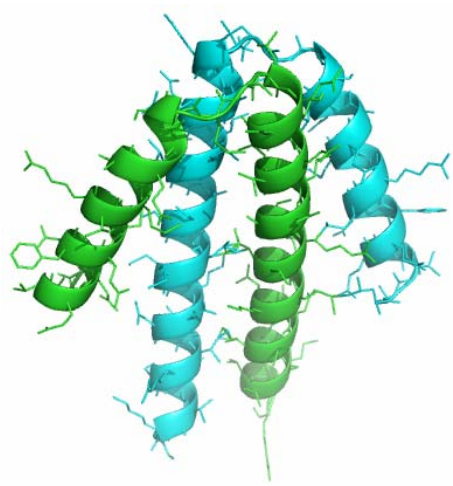


Figure 18. The initial *Tv*-NblA structure obtained after the molecular replacement with Phaser.

The initial R_{factor} of molecular replacement was about 50% but a visual inspection of the electron density maps showed a high correlation between the maps and the model. Many trials of molecular replacement with different initials parameters were done. The best result obtained from molecular replacement with an R_{factor} of about 41% was used for an additional refinement with the program CNS (Brunger et al., 1998). In the beginning we used Rigid-body refinement followed by rounds of positional and temperature factor refinements as well as simulated annealing. Throughout the refinement process we used all data from 20 to 2.5Å. Correction of the model as well as manual rebuilding of loops and side chain orientations was guided by XtalView and Coot corrected $2F_o$-F_c electron density maps. All these refinement procedures allow decreasing the R/R_{free} values at 0.23/0.24. The correlation between the model and the electron density maps can be observed in the figure 19.

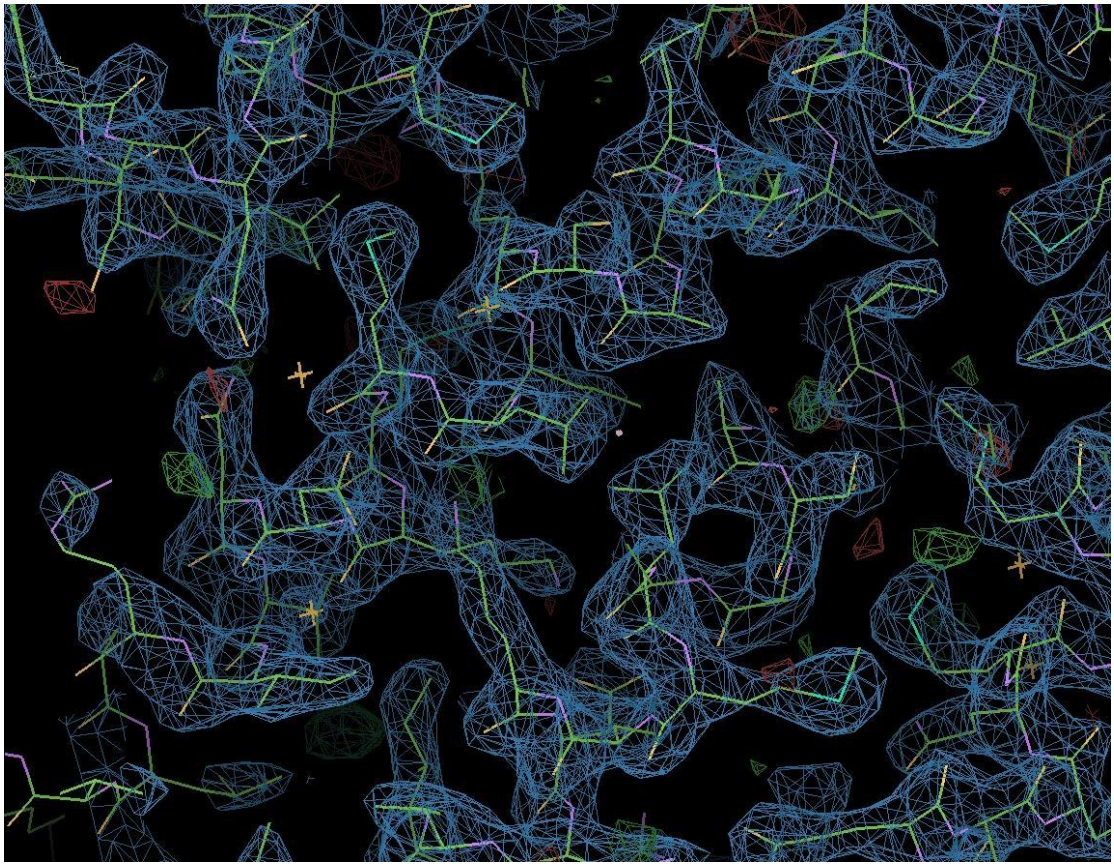


Figure 19. Correlation between the *Tv*-NblA final model and the electron density map.

The quality of the model, reflection data, and structure refinement are presented in Table 3.

Table 3. *Tv*-NblA reflection data and structure refinement statistics

Resolution range (Å)	20-2.5
Number of reflections:	
Total	5328
Working set	4755
Test set	573
Model composition	
Amino acids:	
Total	105
Subunit A	54
Subunit B	51
Total atoms	878
Water molecules:	28
Rms deviation from identity[2]	
Bond distance (Å)	0.018
Angles (^{o})	1.181
Mean B Value (Å^2)	36.43

[1]$R=\Sigma(F_o - F_c)/ \Sigma(F_o)$

6.7.3 The overall structure of *Tv*-NblA

The final model of *Tv*-NblA includes a dimer in the asymmetric unit (see figure 20).

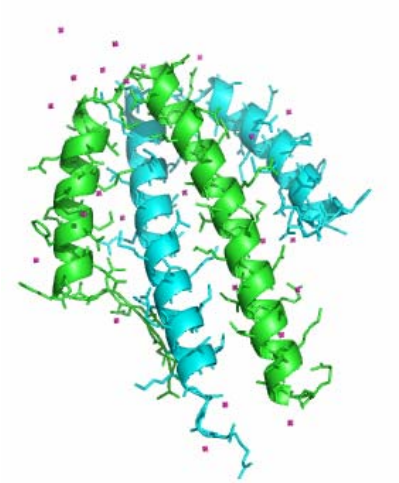


Figure 20. The organization of *Tv*-NblA subunits in the asymmetric unit.

The two copies of the molecule in the crystallographic asymmetric unit do not differ significantly from each other (see figure 21). The r.m.s. deviation value is only 0.433. The principal difference between the two chains is seen especially in the flexible loops at the end of the α-helixes.

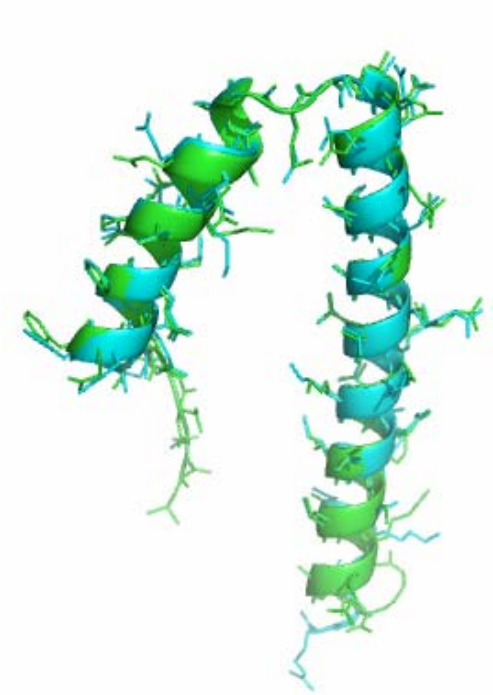


Figure 21. The superposition of the chains of *Tv*-NblA from the same asymmetric unit.

Subunit A and B are well ordered, both of them are formed from two α-helices coupled by a short loop (see figure 20). The angle between the α-helixes is the same for the two subunits. The subunits are complete from the residue 8-62 and 11-62 respectively of the 62-residue native protein. We were also able to define 28 molecules of water in the asymmetric unit (see figure 20).

The overall structure of *Tv*-NblA is similar with the dimer from the *Anabaena*-NblA structure (see figure 22). The long helices are more similar then the short helices. We were able to extend the flexible loops past those built for the *Anabaena* NblA.

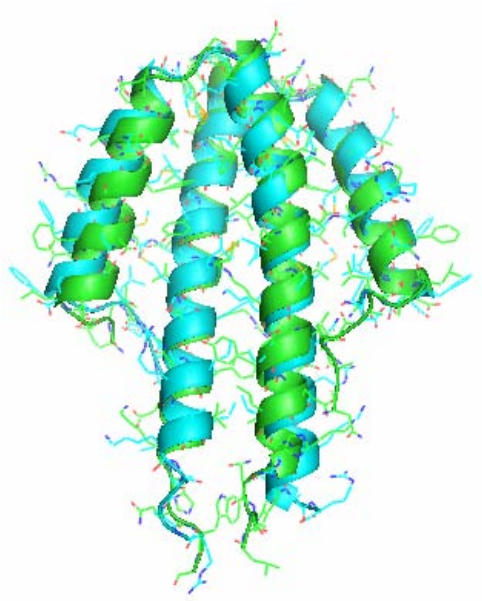
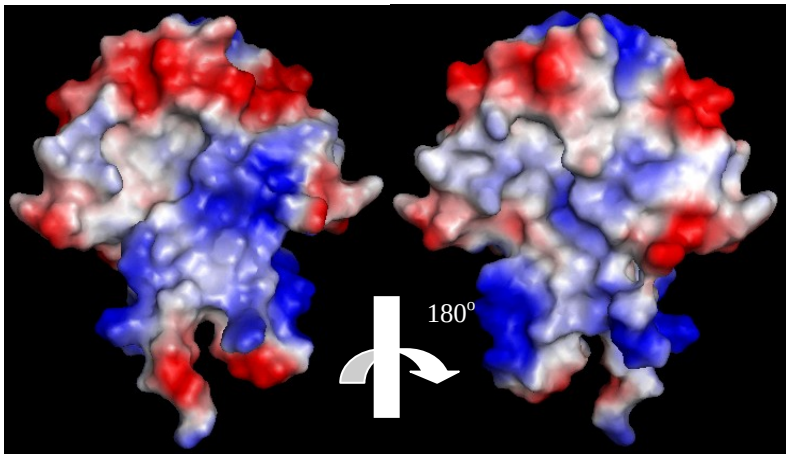


Figure 22. Superposition of the *Tv*-NblA (green) and *Anabaena* NblA (blue).

The calculated electrostatic potential of the surface the *Tv*-NblA and *Anabaena* NblA are very different (see figure 23). *Tv*-NblA has a region with high electronegative potential in the region of the connection loop and the beginning of the α-helices and a region with a high electropositive potential in the middle of the protein in both sides of the molecule. The *Anabaena* NblA is more neutral with an electronegative region in the end of the long α-helixes. This electronegative region is present also in the *Tv*-NblA but is smaller.

The fact that the structure of the NblA proteins is very similar, but their sequence and electrostatic potential are different brings us to suggest that the NblA function is performed mainly due to their structural architecture.



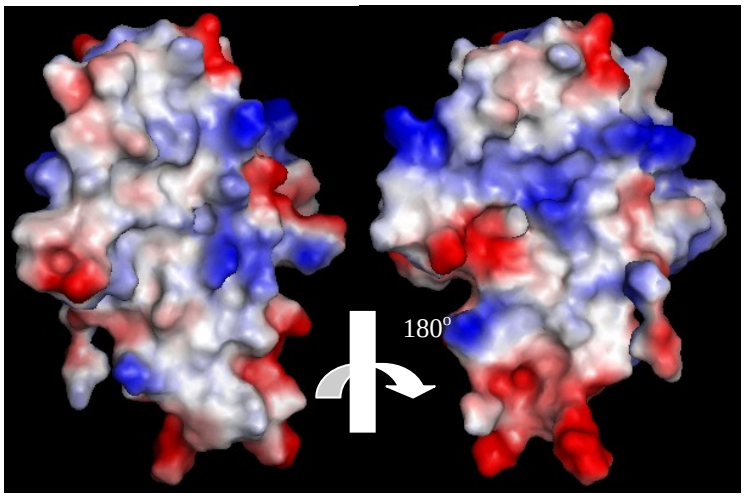
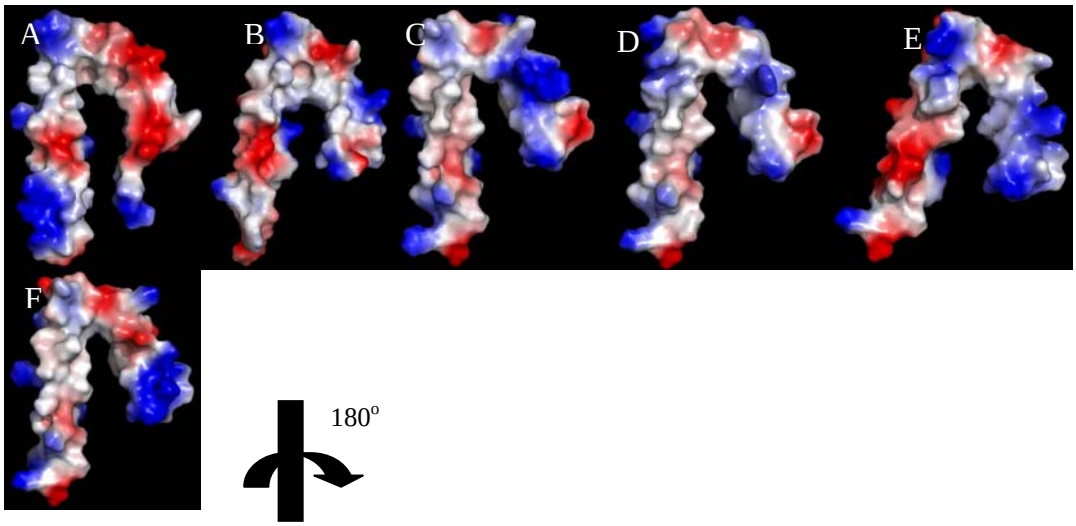


Figure 23. Calculated electrostatic potential of *Tv*-NblA (upper) and *Anabaena* (lower) in vacuum by PyMol (red is electronegative and blue is electropositive).

To a better understand the structure-function relation we modelled the structure of other NblA proteins from different organisms. The models were done using the *Tv* and *Anabaena* NblA structures. We checked the electrostatic potential of all the NblA models in order to observe a region with a consistent potential. It is possible to observe that one of the sides has a constantly electropositive region in the middle of the long α-helices. This region has a correspondent electronegative region on the other side of the protein. Unfortunately, we were not able to observe a more specific or pronounced similarity between the different NblA molecules at this level of information (see figure 24).



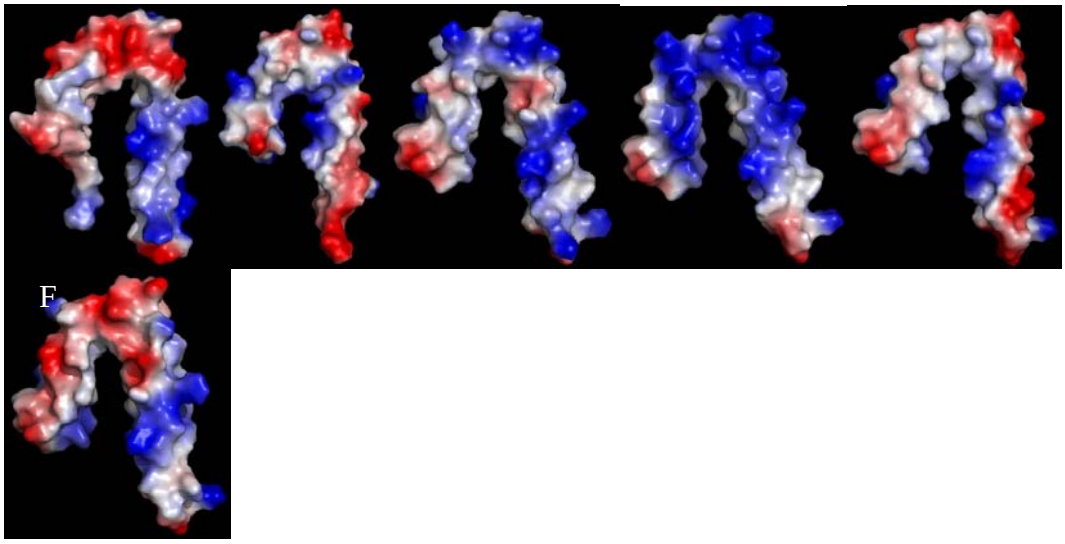


Figure 24. Calculated electrostatic potential in vavum (by PyMol) (red is electronegative and blue is electropositive) comparation between the models of NblA from different organisms from both sides. A) *Tv*-NblA; B) *Anabaena* NblA; C) *Cyanidioschyzon* NblA; D) *Gracillaria* E) ATCC29413 NblA; F) *F. diplosiphon* NblA.

6.7.4 NblA and Phycocyanin Interaction

It is well already all recognized that the NblA molecule is a critical component for the phycobilisome disassembly during nutrient stress. Sequence homology comparisons between the NblA molecules from different photosynthetic species show a low degree of homology (<30% identity) (see figure 3), which is somewhat surprising since the putative substrates (PE, PC or APC) are highly homologous (typically 70-90%). In addition, the NblA has no known homology to any other protein family in the various gene banks, suggesting that its functional activity may be unique. The lack of homology between the NblA proteins hints that this family of proteins have a structure related function with a pure relation to the amino acid sequence. Because NblA consists of two α-helices and the putative substrates (PE, PC or APC) are composed also from α-helices we checked the similarity between the different pairs of α-helices from the PC and the NblA (see figure 25).

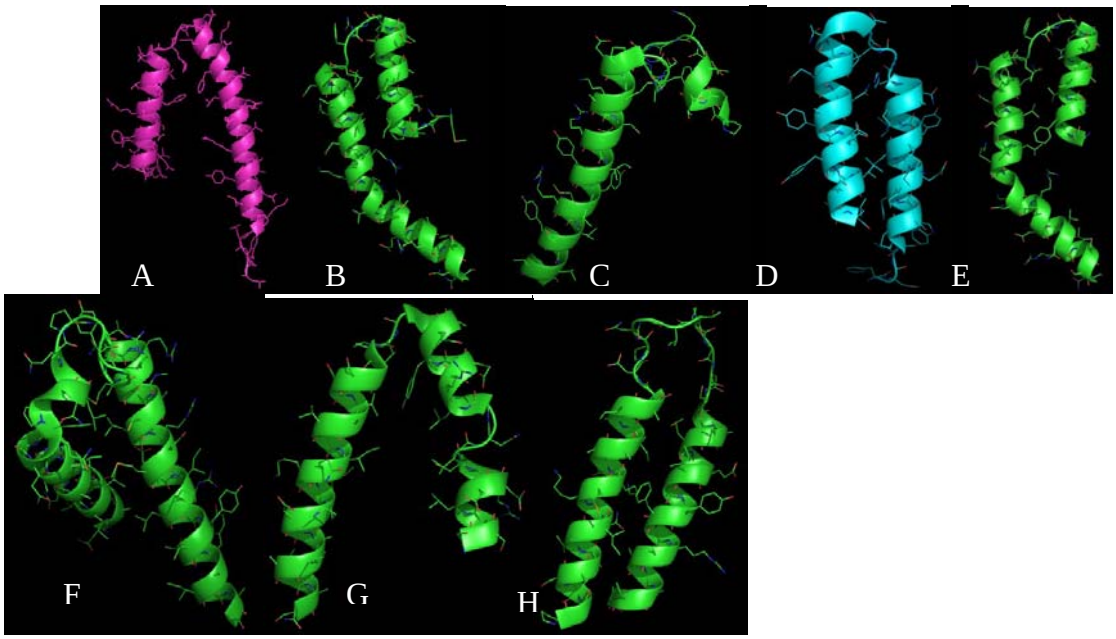


Figure 25. Comparison between the α-helices pairs between the NblA and PC molecules. A) *Tv*-NblA; B-H) Different pairs of α-helices from the *Tv*-PC.

The pair B is formed from the α-helices X, Y and A from the subunit α (from amino acid 122-171/ subunit α). The pair C is formed from the α-helicex B from the subunit α (from amino acid 64-101/ subunit α). The pair D is formed from the α-helices G and H from the subunit α (from amino acid 122-171/ subunit α). The pair E is formed from the α-helices X, Y and A (from amino acid 3-46/ subunit β) from the subunit β. The pair F is formed from the α-helices B and E (from amino acid 47-101/ subunit β) from the subunit β. The pair G is formed from the α-helices F and G from the subunit β (from amino acid 104-145/ subunit β). The pair H is formed from the α-helices G and H (from amino acid 125-172/ subunit β). The names of the α-helices were previously described by Adir et al., 2006.

It is possible to observe that there is some similarity between the NblA and the pair C (RMS = 1.623Å) and G (RMS = 0.486Å) of α-helices from the PC (see figure 25). The similarity is especially in the angle between the two α-helices and in the length of the long α-helix. We were interested also to check the similarity between the electrostatic potential of the NblA and the different pairs of α-helices from the PC (see figure 26).

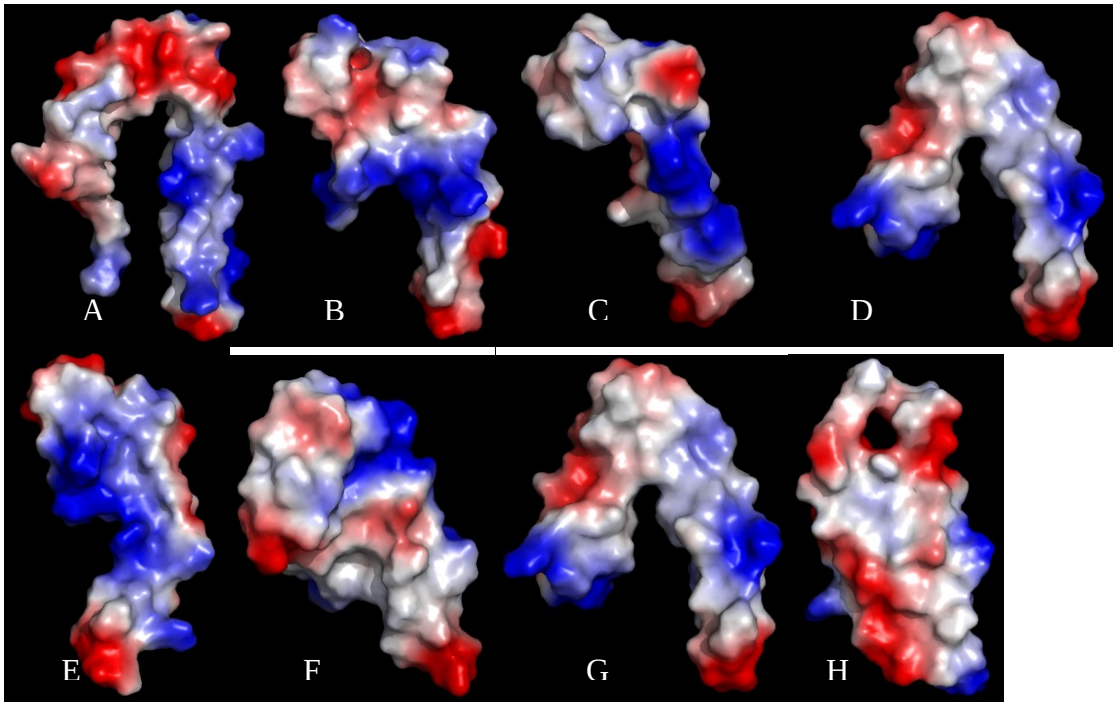


Figure 26. Comparison between the electrostatic potential in vacuum (by PyMol) (red is electronegative and blue is electropositive) of the NblA and the α-helices pairs of the PC molecules. A) *Tv*-NblA electrostatic potential; B-H) Different pairs of α-helices from the *Tv*-PC electrostatic potential.

The similarity between the NblA and pairs C and G of α-helixes from the PC is somewhat consistent also on the electrostatic potential level. The long helix of NblA is formed by a small region with electronegative potential in the beginning, followed by a longer region with electropositive potential and again electronegative potential region at the end of the helix. A similar pattern is also seen in C and G (see figure 26). This similarity is not completely perfect but we think that the NblA is supposed to mimic the C or G pairs from PC and to interfere with the phycobilisome self association only during starvation. This means that a lesser similarity is desirable so that the interaction may be transitive.

6.7.5 Model of NblA-Phycocyanin Interaction

It was already shown that NblA binds to phycobiliproteins (Bienert et al., 2006; Luque et al., 2003). We tried to reproduce these results by many biochemical techniques. The first biochemical technique that we tried was protein affinity chromatography. In this method a protein is coupled to a matrix under controlled conditions and used to select ligand proteins that bind to the protein of interest. If the

two proteins do not associate the ligand protein will pass through the column. If the proteins associate the ligand protein will be retained on the column and will elute together with the protein of interest. In our case we coupled the His-tag-NblA molecules to a Ni column. After the protein was bound to the column we added phycocyanin and washed the column. We found, under a variety of conditions, that at this stage the PC was washed from the column and did not bind to the NblA. This disagrees the previous results that determined interaction between the NblA and the PC (Bienert et al., 2006; Luque et al., 2003).

A different method used to try and determine the interaction between NblA and PC was ITC. In this method we titrated the NblA with trimeric PC and checked the thermal changes. Unfortunately no thermal changes, that can suggest an interaction between the NblA and the PC, were observed. It is possible that the NblA interacts only with larger subunits, such a hexameric PC or rods, or with the whole phycobilisome. To preserve these larger phycobilisome complexes after the breaking of the cyanobacteria cells is more difficult and requires high concentrations of phosphate salt, making the experiments more difficult.

On the basis of the phycocyanin and NblA structures we designed a theoretical model of interaction between the NblA and the phycobilisome (see figure 27).

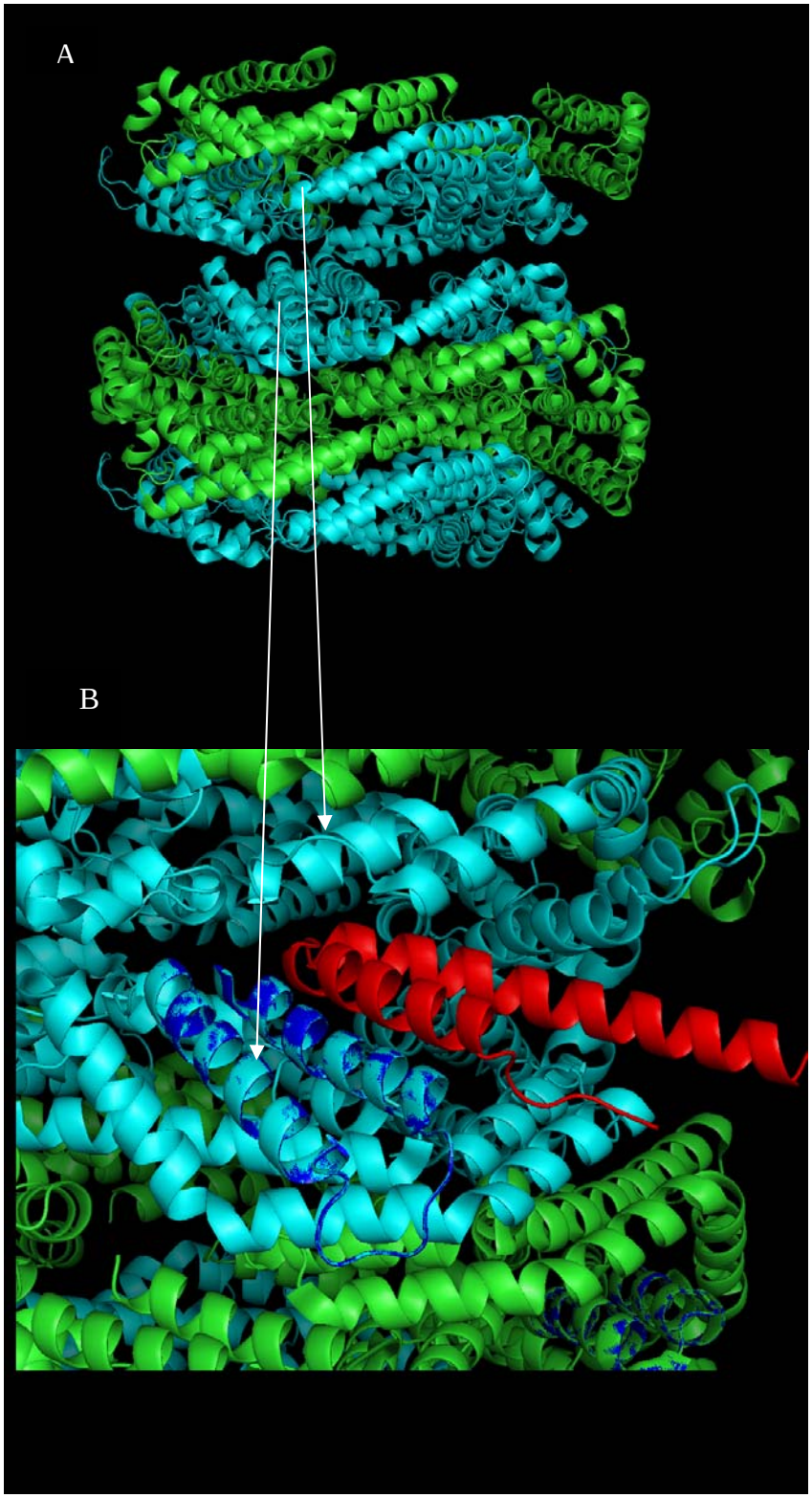


Figure 27. Theoretical model of interaction between the NblA and the phycobilisome. A) Phycobilisome theoretical model. B) NblA-phycobilisome possible interaction.

A possible mode of interaction between the NblA and the phycobilisome is the insertion of the NblA into the phycobilisome during starvation. This insertion may force the disassembly of the phycobilisome. The association of the NblA to the

phycocyanin probably disables the connection of adjacent phycocyanin disks and inhibits the rebuilding of the phycobilisome. In our model, as is possible to see in figure 27, NblA inserts into the phycobilisome near the α-helical H pair (see figure 27B) in a horizontal way. This is the only place in the phycobilisome model where the NblA has enough space to penetrate. From the theoretical model of the phycobilisome (see figure 27A) it is possible to observe that there is not enough space for the NblA to interact with the PC in a vertical position as was previously described by Bienert et al (Bienert et al, 2006) (see figure 5).

The H pair from the PC, where the NblA inserts into the phycobilisome in our theoretical model, has strong interactions with pair C (see figure 28, pair H is coloured in red-orange and pair C is coloured in green-yellow).

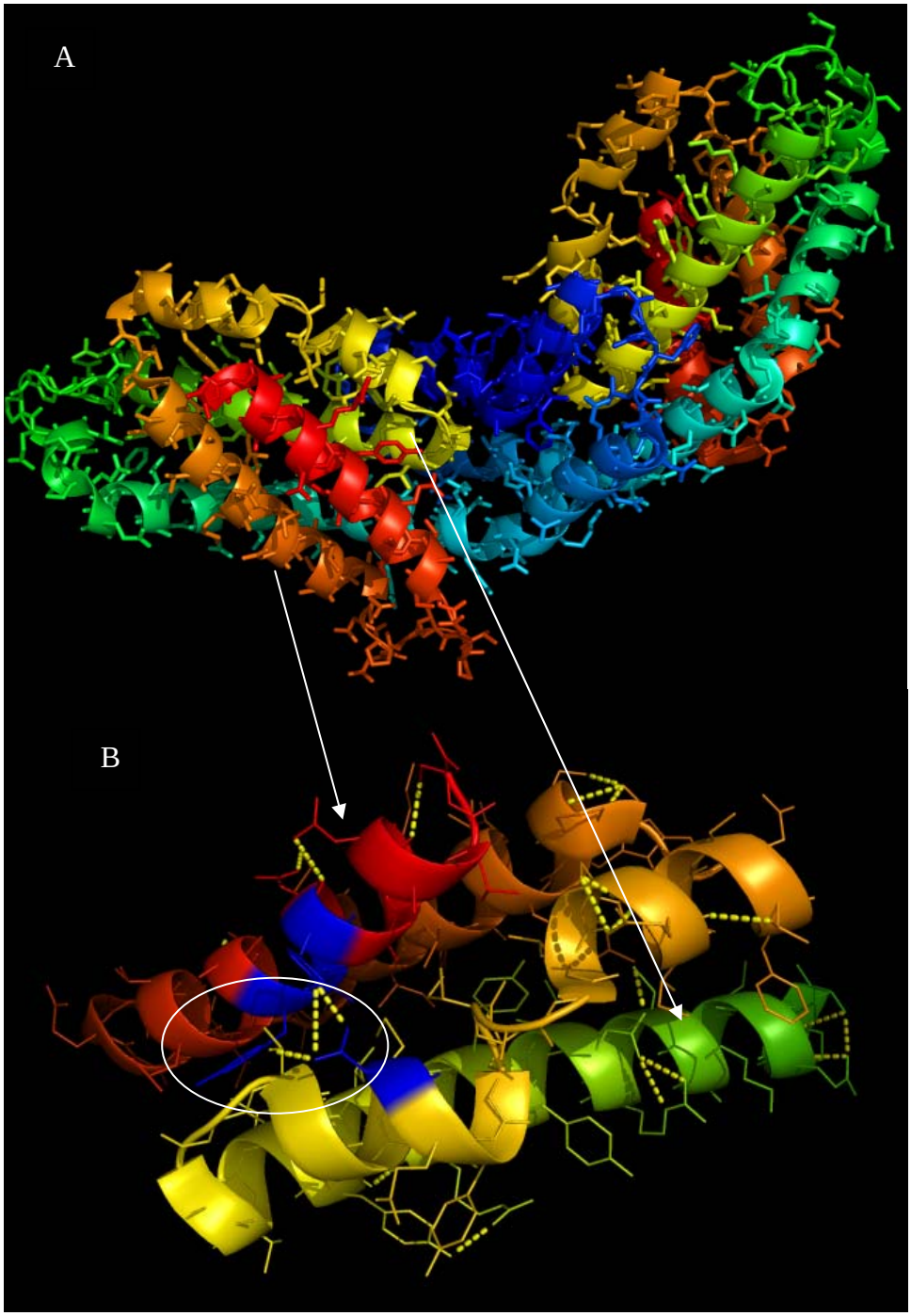


Figure 28. A) PC structure (1KTP) with the interactions between the α-helices pairs B) Interaction between the pair H and C.

As we previously showed, the NblA has some similarity with the pair C, likeness in the angle between the two α-helices and in the electrostatic potential.

A possible model of the NblA function can be related with its structure and similarity to the pair C from the PC. In case of nutrient stress the NblA expression increases probably to a level similar to the PC level. We intend to propose a possible mechanism of function of the NblA that probably mimics the pair C and will connect

to the phycobilisome. This connection may disable the phycobilisome formation or may dissemble the phycobilisome by frustrating the interactions between the different PC α-helices. Further we suggest that during continued starvation the phycobilisome associated NblA is proposed to inhibit the reassociation of the PC subunits.

6.7.6 Structure Determination of *Tv*-NblA without urea

In our many trails to obtain crystals of *Tv*-NblA only once we were able to obtain native *Tv*-NblA without urea. From the tens of crystals obtained in this prep only one diffracted. However, from this one crystal we succeeded to obtain the structure of *Tv*-NblA without urea by molecular replacement. The search model used for the molecular replacement was the structure of *Tv*-NblA dimer with urea. Phasing by molecular replacement was performed, by the program Phaser (Read, 2004) from the CCP4 suit of programs (Collaborative Computational Project, Number 4, 1994). The molecular replacement gave a very high LLG of more then 2000. The R-factor of the first model of *Tv*-NblA without urea was about 30%. The refinement was done by Refmac5 (Murshodov, 1997) from the CCP4 suit of programs (Collaborative Computational Project, Number 4, 1994) and the water molecules were obtained with Coot. At present the R/R_{free} values are 0.25/0.29. The quality of the model, reflection data, and structure refinement are present in Table 4.

Table 4. Statistics of reflection data and structure refinement of the *Tv*-NblA without urea

Resolution range (Å)	47-2.5
Number of reflections:	
Total	9109
Working set	8500
Test set	465
Model composition	
Amino acids:	
Total	210
Subunit A	54
Subunit B	51
Subunit C	54
Subunit D	51
Total atoms	1744
Water molecules:	30
Rms deviation from identity[2]	
Bond distance (Å)	0.036
Angles (^{o})	2.9
Mean B Value (Å^2)	42.58

The overall structure of *Tv*-NblA without urea

The final model of *Tv*-NblA without urea includes a tetramer in the asymmetric unit (figure 29). The tetramer is formed from two almost identical dimers and each dimer is very similar to the dimer of the *Tv*-NblA crystallized in the presence of urea (see figure 30). The r.m.s. deviation value is only 0.18Å.

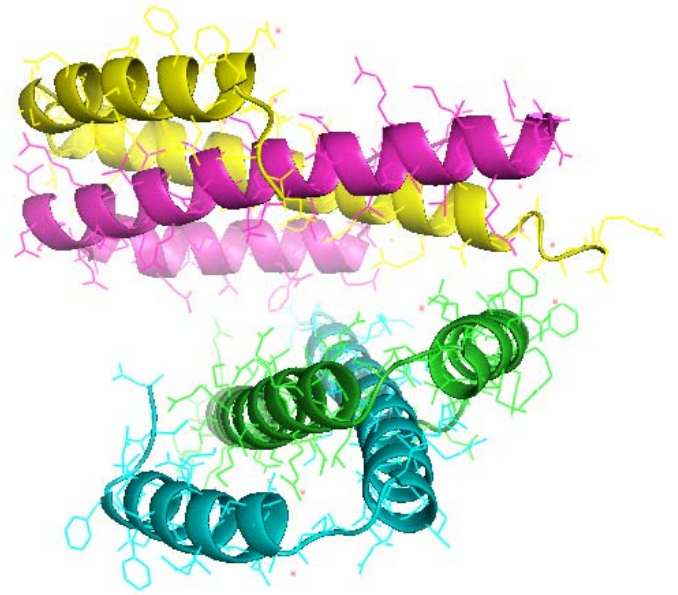


Figure 29. Final structure of *Tv*-NblA without urea.

The similarity between the *Tv*-NblA structures with and without urea demonstrates again the ability of 2M urea to afford solublization without denaturation. This is indeed important in light of the structural proteomics efforts going on world-wide. Proteins with limited solubility are typically discarded in efforts such as these, reducing the number of potential targets (Pusey *et al.*, 2005). It is certainly possible that for other proteins, either higher or lower concentrations of urea may be required; however it seems likely that such a concentration can be found empirically.

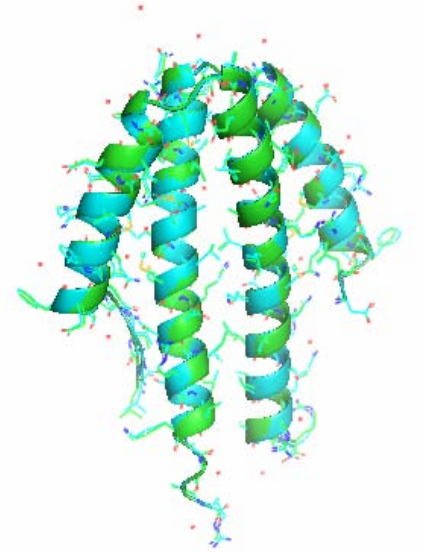


Figure 30. The superposition of the *Tv*-NblA without urea in (P1, green) and the *Tv*-NblA with urea (P3221, blue).

The interaction between the two dimers from the asymmetric unit of the crystals of *Tv*-NblA without urea is limited. The only interaction is between the Glycine 26 and the Arginine 59 (see figure 31).

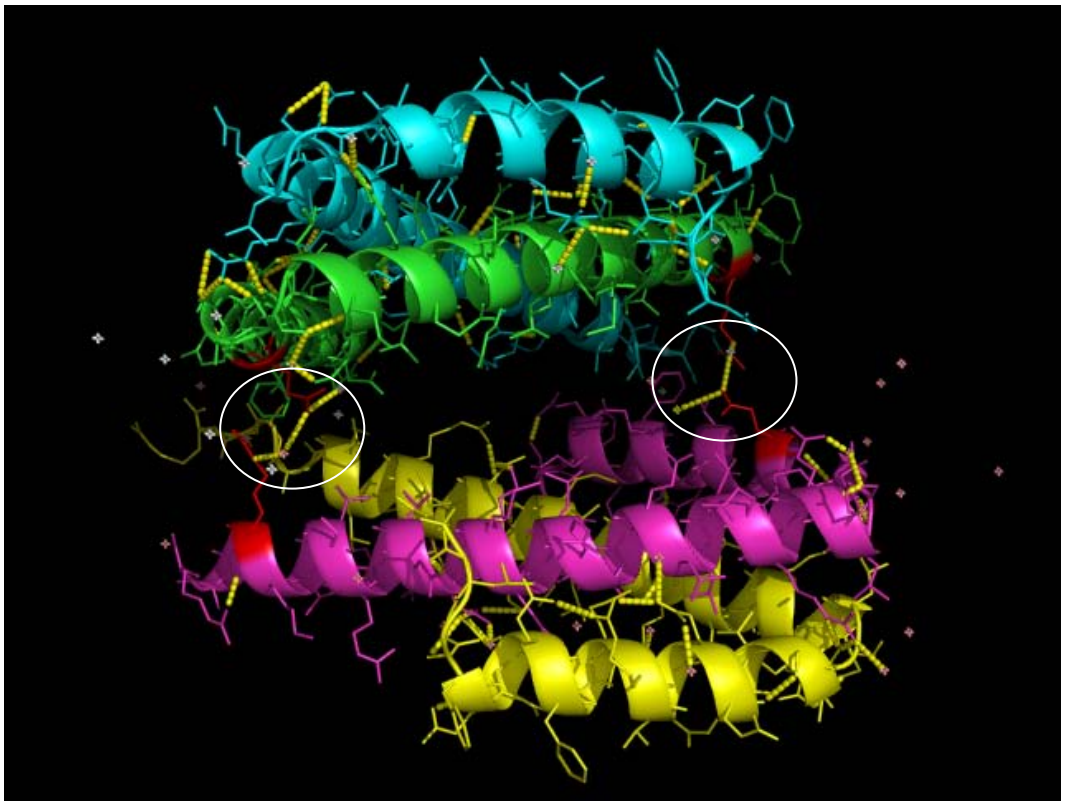


Figure 31. *Tv*-NblA without urea structure. The interaction between the different dimers is presented in red.

6.7.7 Structure Determination of *Se*-NblA

Our many crystallization trials using Selenium *Se*-NblA did not result in the growth of crystals. Because of this our only mode to determine the *Se*-NblA structure was only by molecular replacement. The molecular structure of NblA from *Anabaena* sp. PCC 7120 (Bienert et al., 2006) (PDB code 1OJH) and *Tv*-NblA were used as a search model for molecular replacement. For the molecular replacement we used the one of the dimers of the NblA from *Anabaena* or the dimer from the *Tv*-NblA as a basic model. The level of homology on the amino acid sequence level between the Anabaena NblA or *Tv*-NblA and *Se*-NblA is also poor, about 20% identity. Most of the homologous residues are spread out over most of the protein sequence, and not clustered in certain regions. The final structure used as the additional search model was between the residues 7-54 and was formed only by two α-helices coupled by a short loop for bought of the monomers.

Phasing by molecular replacement was performed, like in the case of *Tv*-NblA, by the program Phaser (Read, 2004) from the CCP4 suit of programs (Collaborative Computational Project, Number 4, 1994). The molecular replacement gives the best LLG of 57 score using a data subset between 20 Å and 2.5Å resolution. The RMS parameter between the two molecules used for the molecular replacement was 1.2.

Using Phaser we were able to obtain an initial *Se*-NblA model formed from four dimers (see figure 32).

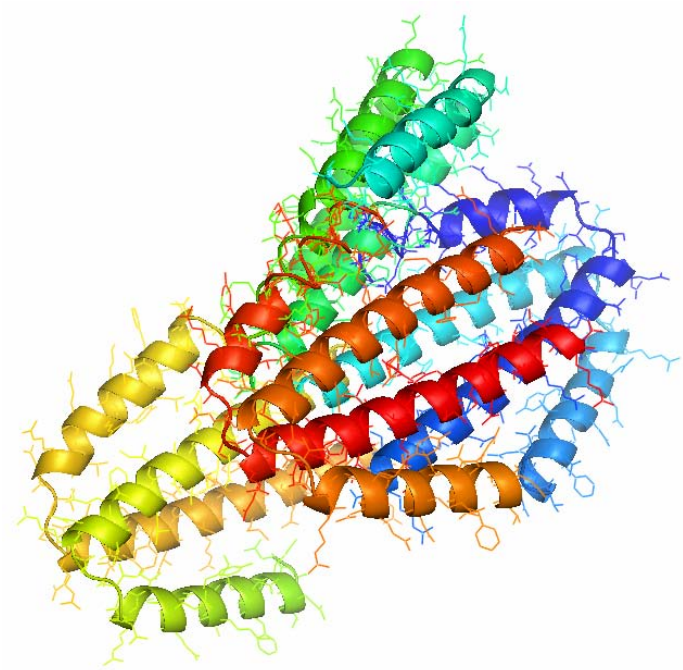


Figure 32. The initial *Se*-NblA model obtained after molecular replacement is an octamer.

The initial R-factor was about 50%. In spite our many refinement trials we were not able to reduce the R-factor. The initial octamer model is very condensed. We decided to delete one of the dimers and to check again the refinement process. In the end, we were able to obtain a more probable *Se*-NblA model formed from three dimers (see figure 33). After many stages of refinement the R-factor of the hexamer *Se*-NblA model is about 40%. As was already described before our data is twinned. The high twinning percentage of the data makes the refinement process very difficult and this is probably the main reason for a relatively high R-factor. Attempts to perform MR using other possible space groups failed completely.

Figure 33. Se-NblA model with an R-factor of about 40% is formed from three dimers.

Dr. Rakefet Schwarz and co-workers performed random point mutations on the whole *Se*-NblA sequence, and screened for non-bleaching mutants. Six of the point mutations appear to affect the bleaching process. Those mutations are located all over the NblA sequence (see figure 34A) which conflicts with the previous results obtained by Bienert *et al.* (Bienert et al., 2006) where the mutations that prevent the bleaching are located at the end of the α-helices (see figure 5).

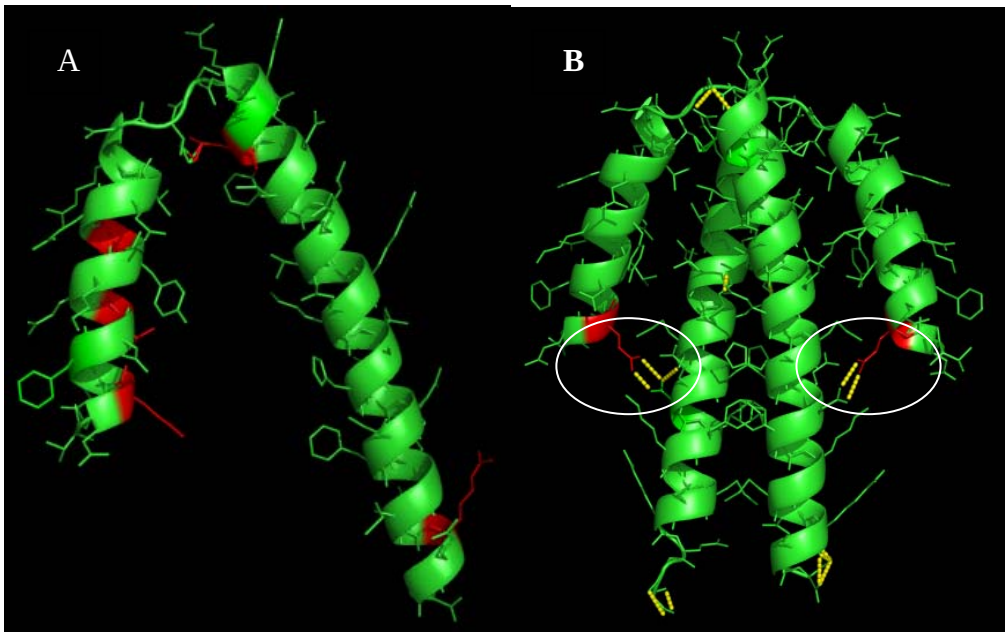


Figure 34. A) *Se*-NblA preliminary model structure with the point mutations that prevent the bleaching process (red). B) *Se* –NblA dimer preliminary model with the point mutations that prevent the bleaching process and are related in the monomers interaction (red).

As can be observed in figure 34 B Glutamine15, which involved in the bleaching process, is involved in the interaction between the two monomers of the dimer. A possible explanation for this is that after expression, the NblA requires the formation of a dimeric structure in order to maintain stability of the protein. This explination is strengthened by the fact that in all the crystal structures (also in those with urea) the NblA was obtained as a dimer. Our hypothesis is that during starvation, the NblA may be disassembled from the dimer into monomers that are able to insert into the phycobilisome.

6.7.8 Structure Determination of HspA

Until now we were able to grow only native crystals of HspA in the presence of urea. For this reason we tried to determine the HspA structure only by molecular replacement. From the four sHSPs structures solved until now we chose the wheat HSP 16.9 as a search model for the molecular replacement. The wheat HSP 16.9 assembles into a dodecameric double disk (figure 8). Each disk is organized as a trimer of dimers. From the huge wheat HSP 16.9 complex we picked only a monomer. The level of homology on the amino acid sequence level between the HspA and wheat

HSP 16.9 is about 20%. Most of the homologous residues are clustered in the α-crystalline region (see figure 9). A short segment that contain only the α-crystalline region (from amino acid 45 to 137) and have a homology of 28% was used as the additional search model (see figure 31).

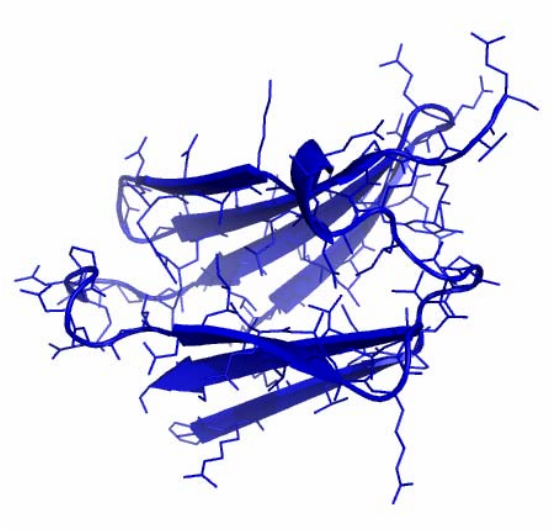


Figure 35. Homology based molecular model of the HspA protein folded according to the structure of wheat HSP 16.9.

Phasing by molecular replacement was performed using the programs Phaser (Read, 2004), AMoRe (Navaza, 1994), MolRep from the CCP4 suit of programs (Collaborative Computational Project, Number 4, 1994) and, also, CNS (Brunger et al., 1998). A few possible solutions with an initial R_{factor} of about 50% were obtained. Attempts to improve the R_{factor} and redefine the model also by CNS and Refmac5 have not yet been successful.

In spite the relatively high similarity between the α-crystalline region of the HspA and wheat HSP 16.9 and our many trails with different programs until now we were not able to obtain the structure of the HspA by molecular replacement. The main reason for this can be the possibility that the β-strands from the HspA are organized in a different mode than those of wheat HSP 16.9. The α-crystalline domain from all the sHSPs solved until now is β -sandwich structures with a seven-strand arrangement. From the bioinformatics predictions of the secondary structure that we done with different programs (http://www.predictprotein.org, http://www.compbio.dundee.ac.uk/~www-jpred) results that the α-crystalline domain from the HspA is formed also from seven β-strands. It is possible that the angles

between the strands are different in the case of HspA and this make the molecular replacement more difficult.

Another possibility for our lack of success in solving the HspA structure by molecular replacement can be the use of urea in the crystallization conditions. In spite of the results shown with the *Tv*-NblA, there still exists the possibility that the relatively high concentration of urea (2M), influences the structural conformation of the α-crystalline domain from the HspA.

7 Conclusions and plans for future research

Under conditions of nutrient starvation, expression of the NblA protein is required in the process that leads to phycobilisome degradation and bleaching of the cells. HspA, a 16.5kDa heat shock protein expressed in cyanobacterial cells, has been shown to provide functional stability to the phycobilisome during heat stress. In the work presented in this thesis, we have cloned the genes encoding for the *Tv*-NblA, *Se*-NblA and HspA proteins into bacterial expression vectors in order to determine their three-dimensional structures. The resulting recombinant proteins were found to be sparingly soluble, limiting their usefulness in the performance of crystallization experiments. We have developed a novel protocol that utilizes relatively high concentrations of urea to afford sufficient solubility to the protein. This has lead to the successful growth of diffraction quality crystals of these proteins. Complete data sets collected to 2-2.5Å from crystals of both proteins shows that the crystals are stable, and useful for structure determination.

The mode of interaction between urea and soluble proteins has been extensively studied (Bolen, 2004; Bolen and Baskakov, 2001; Street et al., 2006). The denaturing effect of urea has been proposed to be primarily a result of the association of urea molecules with the backbone of proteins, thereby forcing typically buried residues to interact with the solvent. At relatively low concentrations, urea can apparently serve as a solubilizing agent, which can even lead to crystal formation (Bolen, 2004). In this report, we show that even higher urea concentrations do not hinder the crystallization process. It is possible, that at relatively high protein concentrations, the urea merely prevents unspecific aggregation, reducing these proteins tendency to precipitate, but not nucleate.

Interestingly, the ability of 2M urea to afford solublization without denaturation was not secondary structure dependent, since the NblA is a wholly α-helical protein, while members of the sHSP family have been shown to contain extensive β-sheets (Dgany et al., 2004; Kim et al., 1998; Stamler et al., 2005; van Montfort et al., 2001). The fact that crystals were readily obtained, and the fact that the *Tv*-NblA with urea structure determined shows no denaturative effects and is

similar to the structure of *Tv*-NblA without urea that was also determined, suggests that this method can be used routinely in the process of crystallization trials.

In the work presented in this thesis we succeeded to determine the structure of the NblA protein from the thermophilic cyanobacteria *T. vulcanus* at 2.5Å by molecular replacement. The NblA protein has been shown to be physically involved in the disassembly of the phycobilisome antenna complex in cyanobacteria and red-algae during nutrient starvation. The monomer has a helix-hairpin-helix motif which dimerizes into an open four-helical bundle in the asymmetric unit, identical to the NblA structure from *Anabena*. The major experimental tool used was X-ray crystallography. Additional biochemical methods were utilized to try and confirm the NblA –PC interaction demonstrated in previous studies (Bienert et al., 2006) without any positive result. One of our future aims is to study the mode of interaction between the NblA proteins and the phycobilisome. For this aim we proposed other *in vitro* biochemical studies that are supposed to check the affinity of the NblA to the complete phycobilisome. We suggest mutagenesis of the NblA protein that is supposed to show the critical residues that affect the disassembly process (bleaching) *in vivo*. We also intend to obtain a final structure of the *Se*-NblA molecule, and to exchange *in vivo* the genes for NblA proteins in *T. vulcanus* and *S. elongatus* and to check the influence on the bleaching process.

Comparison between structures and homology based models indicate that while the NblA family is only slightly homologous, there are certain attributes of the protein that remain constant, like the protein structure, and could be involved in its function, which is probably a structure related function. The electrostatic potential of the proteins from the NblA family is also dissimilar. We also proposed a computational theoretical model of interaction between the NblA and the phycobilisome. In this theoretical model we show how the NblA molecule can insert into the phycobilesome in case of starvation.

In this work we were able to obtain quality native HspA crystals. Despite our many trails and the many different programs that we used we were not able to solve the phase problem and to obtain the structure by molecular replacement. Future growth of HspA Se-Met derivate crystals will allow us to solve the phase problem by multiple-wavelength anomalous diffraction. We also tried to crystallize and determine the structure of other cyanobacterial proteins from the sHSP family.

Future determination of the structure of the NblA and sHSP from different cyanobacteria will give us the possibility to understand more about the mechanisms by which cyanobacteria deal with different forms of physiological stress. We also hope to understand more about the mode of interaction of the photosystem machinery with other proteins.

Adams, W.W., and C.B. Osmond. 1988. Internal CO(2) Supply during Photosynthesis of Sun and Shade Grown CAM Plants in Relation to Photoinhibition. *Plant Physiol.* 86:117-123.

Adir, N. 2005. Elucidation of the molecular structures of components of the phycobilisome: reconstructing a giant. *Photosynth Res*. 85:15-32.

Adir, N., M. Dines, M. Klartag, A. McGregor, and M. Melamed-Frank. 2006. Assembly and Disassembly of Phycobilisomes. *In* Microbiology Monographs: Inclusions in Prokaryotes. Vol. 2. J.M. Shively, editor. Springer Berlin / Heidelberg. 47-77.

Adir, N., and N. Lerner. 2003. The crystal structure of a novel unmethylated form of C-phycocyanin, a possible connector between cores and rods in pycobilisomes. *J Biol Chem*. 278:25926-32.

Adir, N., R. Vainer, and N. Lerner. 2002. Refined structure of C-phycocyanin from the cyanobacterium Synechococcus vulcanus at 1.6 A: insights into the role of solvent molecules in thermal stability and co-factor structure. *Biochim Biophys Acta*. 1556:168-74.

Adir, N., H. Zer, S. Shochat, and I. Ohad. 2003. Photoinhibition - a historical perspective. *Photosynth Res*. 76:343-370.

Allen, M.M., and A.J. Smith. 1969. Nitrogen chlorosis in blue-green algae. *Arch Mikrobiol*. 69:114-20.

Anderson, L.K., and C.M. Toole. 1998. A model for early events in the assembly pathway of cyanobacterial phycobilisomes. *Mol Microbiol*. 30:467-74.

Arrigo, A.P. 2000. sHsp as novel regulators of programmed cell death and tumorigenicity. *Pathol Biol (Paris).* 48:280-8.

Asayama, M. 2006. Regulatory system for light-responsive gene expression in photosynthesizing bacteria: cis-elements and trans-acting factors in transcription and post-transcription. *Biosci Biotechnol Biochem.* 70:565-73.

Basha, E., G.J. Lee, L.A. Breci, A.C. Hausrath, N.R. Buan, K.C. Giese, and E. Vierling. 2004. The identity of proteins associated with a small heat shock protein during heat stress in vivo indicates that these chaperones protect a wide range of cellular functions. *J Biol Chem.* 279:7566-75. Epub 2003 Dec 8.

Benedek, G.B., J. Pande, G.M. Thurston, and J.I. Clark. 1999. Theoretical and experimental basis for the inhibition of cataract. *Prog Retin Eye Res.* 18:391-402.

Benndorf, R., K. Hayess, S. Ryazantsev, M. Wieske, J. Behlke, and G. Lutsch. 1994. Phosphorylation and supramolecular organization of murine small heat shock protein HSP25 abolish its actin polymerization-inhibiting activity. *J Biol Chem.* 269:20780-4.

Bienert, R., K. Baier, R. Volkmer, W. Lockau, and U. Heinemann. 2006. Crystal structure of NblA from Anabaena sp. PCC 7120, a small protein playing a key role in phycobilisome degradation. *J Biol Chem*. 281:5216-23.

Bolen, D.W. 2004. Effects of naturally occurring osmolytes on protein stability and solubility: issues important in protein crystallization. *Methods*. 34:312-22.

Bolen, D.W., and I.V. Baskakov. 2001. The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J Mol Biol*. 310:955-63.

Bond, U., and M.J. Schlesinger. 1986. The chicken ubiquitin gene contains a heat shock promoter and expresses an unstable mRNA in heat-shocked cells. *Mol Cell Biol.* 6:4602-10.

Brejc, K., R. Ficner, R. Huber, and S. Steinbacher. 1995. Isolation, crystallization, crystal structure analysis and refinement of allophycocyanin from the

cyanobacterium Spirulina platensis at 2.3 A resolution. *J Mol Biol*. 249:424-40.

Brunger, A.T. 1988. Crystallographic refinement by simulated annealing. Application to a 2.8 A resolution structure of aspartate aminotransferase. *J Mol Biol.* 203:803-16.

Brunger, A.T. 1997. X-ray crystallography and NMR reveal complementary views of structure and dynamics. *Nat Struct Biol.* 4:862-5.

Brunger, A.T., P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, and G.L. Warren. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*. 54:905-21.

Canet, D., M. Sunde, A.M. Last, A. Miranker, A. Spencer, C.V. Robinson, and C.M. Dobson. 1999. Mechanistic studies of the folding of human lysozyme and the origin of amyloidogenic behavior in its disease-related variants. *Biochemistry.* 38:6419-27.

Cleveland, D.W., S.G. Fischer, M.W. Kirschner, and U.K. Laemmli. 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J Biol Chem.* 252:1102-6.

Collier, J.L., and A.R. Grossman. 1994. A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *Embo J*. 13:1039-47.

Dgany, O., A. Gonzalez, O. Sofer, W. Wang, G. Zolotnitsky, A. Wolf, Y. Shoham, A. Altman, S.G. Wolf, O. Shoseyov, and O. Almog. 2004. The structural basis of the thermostability of SP1, a novel plant (Populus tremula) boiling stable protein. *J Biol Chem*. 279:51516-23.

Dolganov, N., and A.R. Grossman. 1999. A polypeptide with similarity to phycocyanin alpha-subunit phycocyanobilin lyase involved in degradation of phycobilisomes. *J Bacteriol*. 181:610-7.

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

Ehrnsperger, M., S. Graber, M. Gaestel, and J. Buchner. 1997. Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. *Embo J.* 16:221-9.

Emsley, P., and K. Cowtan. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 60:2126-32. Epub 2004 Nov 26.

Glauser, M., D.A. Bryant, G. Frank, E. Wehrli, S.S. Rusconi, W. Sidler, and H. Zuber. 1992. Phycobilisome structure in the cyanobacteria Mastigocladus laminosus and Anabaena sp. PCC 7120. *Eur J Biochem*. 205:907-15.

Glazer, A.N. 1989. Light guides. Directional energy transfer in a photosynthetic antenna. *J Biol Chem*. 264:1-4.

Grossman, A.R., D. Bhaya, K.E. Apt, and D.M. Kehoe. 1995. Light-harvesting complexes in oxygenic photosynthesis: diversity, control, and evolution. *Annu Rev Genet*. 29:231-88.

Grossman, A.R., M.R. Schaefer, G.G. Chiang, and J.L. Collier. 1993. The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiol Rev*. 57:725-49.

Haley, D.A., J. Horwitz, and P.L. Stewart. 1998a. The small heat-shock protein, alphaB-crystallin, has a variable quaternary structure. *J Mol Biol*. 277:27-35.

Haley, D.A., J. Horwitz, and P.L. Stewart. 1998b. The small heat-shock protein, alphaB-crystallin, has a variable quaternary structure. *J Mol Biol.* 277:27-35.

Haslbeck, M. 2002. sHsps and their role in the chaperone network. *Cell Mol Life Sci*. 59:1649-57.

Haslbeck, M., T. Franzmann, D. Weinfurtner, and J. Buchner. 2005. Some like it hot: the structure and function of small heat-shock proteins. *Nat Struct Mol Biol.* 12:842-6.

Haslbeck, M., S. Walke, T. Stromer, M. Ehrnsperger, H.E. White, S. Chen, H.R. Saibil, and J. Buchner. 1999. Hsp26: a temperature-regulated chaperone. *Embo J.* 18:6744-51.

Hendrickson, W.A. 1991. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science.* 254:51-8.

Hendrickson, W.A., J.L. Smith, R.P. Phizackerley, and E.A. Merritt. 1988. Crystallographic structure analysis of lamprey hemoglobin from anomalous dispersion of synchrotron radiation. *Proteins.* 4:77-88

Irie, T., W. Koyama, Y. Ikeuchi, and T. Kanaseki. 1991. On the three-dimensional structure of quick-frozen hepatic Mallory bodies with special reference to the appearance of cytoplasmic vesicles. *Cell Struct Funct.* 16:1-16.

Jofre, A., M. Molinas, and M. Pla. 2003. A 10-kDa class-CI sHsp protects E. coli from oxidative and high-temperature stress. *Planta.* 217:813-9. Epub 2003 May 13.

Johnston, J.A., C.L. Ward, and R.R. Kopito. 1998. Aggresomes: a cellular response to misfolded proteins. *J Cell Biol.* 143:1883-98.

Jones, T.A., J.Y. Zou, S.W. Cowan, and M. Kjeldgaard. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A.* 47:110-9.

Kanervo, E., M. Suorsa, and E.M. Aro. 2005. Functional flexibility and acclimation of the thylakoid membrane. *Photochem Photobiol Sci*. 4:1072-80.

Kelly, S.M., T.J. Jess, and N.C. Price. 2005. How to study proteins by circular dichroism. *Biochim Biophys Acta*. 1751:119-39.

Khamis, S., T. Lamaze, Y. Lemoine, and C. Foyer. 1990. Adaptation of the Photosynthetic Apparatus in Maize Leaves as a Result of Nitrogen Limitation : Relationships between Electron Transport and Carbon Assimilation. *Plant Physiol.* 94:1436-1443.

Kim, K.K., R. Kim, and S.H. Kim. 1998. Crystal structure of a small heat-shock protein. *Nature*. 394:595-9.

Knauf, U., H. Bielka, and M. Gaestel. 1992. Over-expression of the small heat-shock protein, hsp25, inhibits growth of Ehrlich ascites tumor cells. *FEBS Lett.* 309:297-302.

Laskowska, E., A. Wawrzynow, and A. Taylor. 1996. IbpA and IbpB, the new heat-shock proteins, bind to endogenous Escherichia coli proteins aggregated intracellularly by heat shock. *Biochimie.* 78:117-22.

Lavoie, J.N., E. Hickey, L.A. Weber, and J. Landry. 1993. Modulation of actin microfilament dynamics and fluid phase pinocytosis by phosphorylation of heat shock protein 27. *J Biol Chem.* 268:24210-4.

Lindqvist, Y., and C.I. Branden. 1985. Structure of glycolate oxidase from spinach. *Proc Natl Acad Sci U S A.* 82:6855-6859.

Litt, M., P. Kramer, D.M. LaMorticella, W. Murphey, E.W. Lovrien, and R.G. Weleber. 1998. Autosomal dominant congenital cataract associated with a

missense mutation in the human alpha crystallin gene CRYAA. *Hum Mol Genet.* 7:471-4.

Luque, I., J.A. Ochoa De Alda, C. Richaud, G. Zabulon, J.C. Thomas, and J. Houmard. 2003. The NblAI protein from the filamentous cyanobacterium Tolypothrix PCC 7601: regulation of its expression and interactions with phycobilisome components. *Mol Microbiol*. 50:1043-54.

MacColl, R. 1998. Cyanobacterial phycobilisomes. *J Struct Biol*. 124:311-34.

McRee, D.E., and David, P. R. 1999. Practical Protein Chemistry. *Academic Press*. 2nd Edition.

Monson, R.K., C.H. Jaeger, W.W. Adams, E.M. Driggers, G.M. Silver, and R. Fall. 1992. Relationships among Isoprene Emission Rate, Photosynthesis, and Isoprene Synthase Activity as Influenced by Temperature. *Plant Physiol*. 98:1175-1180.

Muchowski, P.J., and J.I. Clark. 1998. ATP-enhanced molecular chaperone functions of the small heat shock protein human alphaB crystallin. *Proc Natl Acad Sci U S A.* 95:1004-9.

Muchowski, P.J., G.J. Wu, J.J. Liang, E.T. Adman, and J.I. Clark. 1999. Site-directed mutations within the core "alpha-crystallin" domain of the small heat-shock protein, human alphaB-crystallin, decrease molecular chaperone functions. *J Mol Biol.* 289:397-411.

Nakamoto, H., and D. Honma. 2006. Interaction of a small heat shock protein with light-harvesting cyanobacterial phycocyanins under stress conditions. *FEBS Lett*. 580:3029-34.

Navaza, J. 1994. AMoRe: an automated package for molecular replacement. *Acta Cryst.* A50:157-163.

Nicholl, I.D., and R.A. Quinlan. 1994. Chaperone activity of alpha-crystallins modulates intermediate filament assembly. *Embo J.* 13:945-53.

Nowitzki, U., A. Flechner, J. Kellermann, M. Hasegawa, C. Schnarrenberger, and W. Martin. 1998. Eubacterial origin of nuclear genes for chloroplast and cytosolic glucose-6-phosphate isomerase from spinach: sampling eubacterial gene diversity in eukaryotic chromosomes through symbiosis. *Gene.* 214:205-13.

Perng, M.D., L. Cairns, I.P. van den, A. Prescott, A.M. Hutcheson, and R.A. Quinlan. 1999. Intermediate filament interactions can be altered by HSP27 and alphaB-crystallin. *J Cell Sci.* 112:2099-112.

Perutz, M.F., B.J. Pope, D. Owen, E.E. Wanker, and E. Scherzinger. 2002. Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid beta-peptide of amyloid plaques. *Proc Natl Acad Sci U S A.* 99:5596-600.

Petko, L., and S. Lindquist. 1986. Hsp26 is not required for growth at high temperatures, nor for thermotolerance, spore development, or germination. *Cell.* 45:885-94.

Pusey, M.L., Z.J. Liu, W. Tempel, J. Praissman, D. Lin, B.C. Wang, J.A. Gavira, and J.D. Ng. 2005. Life in the fast lane for protein crystallization and X-ray crystallography. *Prog Biophys Mol Biol*. 88:359-86.

Raj Lakshmi, M., H.R. Krishna-Murthy, and T.V. Ramakrishnan. 1988. Density-wave theory of dislocations in crystals. *Physical Review B. Condensed Matter*. 37:1936-1949.

Renkawek, K., C.E. Voorter, G.J. Bosman, F.P. van Workum, and W.W. de Jong. 1994. Expression of alpha B-crystallin in Alzheimer's disease. *Acta Neuropathol (Berl).* 87:155-60.

Richaud, C., G. Zabulon, A. Joder, and J.C. Thomas. 2001. Nitrogen or sulfur starvation differentially affects phycobilisome degradation and expression of the NblA gene in Synechocystis strain PCC 6803. *J Bacteriol*. 183:2989-94.

Rollet, E., J.N. Lavoie, J. Landry, and R.M. Tanguay. 1992. Expression of Drosophila's 27 kDa heat shock protein into rodent cells confers thermal resistance. *Biochem Biophys Res Commun.* 185:116-20.

Rossmann, M.G.B., D. M. 1962. *Acta Cryst.* 15.

Roy, S.K., T. Hiyama, and H. Nakamoto. 1999. Purification and characterization of the 16-kDa heat-shock-responsive protein from the thermophilic cyanobacterium Synechococcus vulcanus, which is an alpha-crystallin-related, small heat shock protein. *Eur J Biochem*. 262:406-16.

Roy, S.K., and H. Nakamoto. 1998. Cloning, characterization, and transcriptional analysis of a gene encoding an alpha-crystallin-related, small heat shock protein from the thermophilic cyanobacterium Synechococcus vulcanus. *J Bacteriol*. 180:3997-4001.

Schwarz, R., and A.R. Grossman. 1998. A response regulator of cyanobacteria integrates diverse environmental signals and is critical for survival under extreme conditions. *Proc Natl Acad Sci U S A*. 95:11008-13.

Sendersky, E., R. Lahmi, J. Shaltiel, A. Perelman, and R. Schwarz. 2005. NblC, a novel component required for pigment degradation during starvation in Synechococcus PCC 7942. *Mol Microbiol*. 58:659-68.

Sidler, W.A. 1994. Phycobilisome and phycobiliprotein structures. *In* The Molecular Biology of Cyanobacteria. D.A. Bryant, editor. Kluwer Academic Publishers, Dordrect. 139-216.

Stamler, R., G. Kappe, W. Boelens, and C. Slingsby. 2005. Wrapping the alpha-crystallin domain fold in a chaperone assembly. *J Mol Biol*. 353:68-79.

Street, T.O., D.W. Bolen, and G.D. Rose. 2006. A molecular mechanism for osmolyte-induced protein stability. *Proc Natl Acad Sci U S A.* 103:13997-4002.

Stromer, T., E. Fischer, K. Richter, M. Haslbeck, and J. Buchner. 2004. Analysis of the regulation of the molecular chaperone Hsp26 by temperature-induced dissociation: the N-terminal domail is important for oligomer assembly and the binding of unfolding proteins. *J Biol Chem.* 279:11222-8. Epub 2004 Jan 13.

Susek, R.E., and S.L. Lindquist. 1989. hsp26 of Saccharomyces cerevisiae is related to the superfamily of small heat shock proteins but is without a demonstrable function. *Mol Cell Biol.* 9:5265-71.

Thomas, J.G., and F. Baneyx. 1998. Roles of the Escherichia coli small heat shock proteins IbpA and IbpB in thermal stress management: comparison with ClpA, ClpB, and HtpG In vivo. *J Bacteriol.* 180:5165-72.

Torok, Z., P. Goloubinoff, I. Horvath, N.M. Tsvetkova, A. Glatz, G. Balogh, V. Varvasovszki, D.A. Los, E. Vierling, J.H. Crowe, and L. Vigh. 2001. Synechocystis HSP17 is an amphitropic protein that stabilizes heat-stressed membranes and binds denatured proteins for subsequent chaperone-mediated refolding. *Proc Natl Acad Sci U S A.* 98:3098-103. Epub 2001 Feb 27.

van Montfort, R.L., E. Basha, K.L. Friedrich, C. Slingsby, and E. Vierling. 2001. Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat Struct Biol*. 8:1025-30.

van Waasbergen, L.G., N. Dolganov, and A.R. Grossman. 2002. nblS, a gene involved in controlling photosynthesis-related gene expression during high

light and nutrient stress in Synechococcus elongatus PCC 7942. *J Bacteriol*. 184:2481-90.

Wisniewski, T., and J.E. Goldman. 1998. Alpha B-crystallin is associated with intermediate filaments in astrocytoma cells. *Neurochem Res.* 23:385-92.

Xiao, G., W. Chen, and R.J. Kulmacz. 1999. Comparison of prostaglandin H synthase-1 and -2 structural stabilities. *Adv Exp Med Biol.* 469:115-8.:27:22.

Yokoyama, S. 2003. Protein expression systems for structural genomics and proteomics. *Curr Opin Chem Biol*. 7:39-43.

Zhu, Y., S. O'Neill, J. Saklatvala, L. Tassi, and M.E. Mendelsohn. 1994. Phosphorylated HSP27 associates with the activation-dependent cytoskeleton in human platelets. *Blood.* 84:3715-23.

תקציר

פיקובילוזום הוא קומפלקס אנטנות פוטוסינטטוית. הנמצא בציאנובקטריאות ובאצות האדומות. לפיקוביליזומים יש אפשרות לעבור אינטרקציה עם חלבונים שעוברים ביטוי במיוחד כדי לאפשר לציאנובקטריות להתמודד עם צורות שונות של סטרס פיסיולוגי.

דרוש NblA עולה. חלבון NblAבמקרה של תנאי הרעבה הביטוי של החלבון בתהליך הגורם לפירוק פיקוביליזום וכתוצאה מכך ציאנובקטריות משנות את צבעם מירוק לצהוב בתהליך הנקרא הלבנת (bleaching).

מבוטא יתר כתוצאה משוק תרמי. נמצא כי יש לו 16.5kDa חלבון בגודל HspA תפקיד בשמירה על הפיקוביליזום במקרה של עקת חום.

הצלחנו לשבט את הגנים המקודדים לחלבונים האלו לתוך וקטורי ביטוי חיידקיים, על מנת לקבוע את המבנה התלת מימדי שלהם. החלבונים הרקומבינטיים שהתקבלו התאפיינו במסיסות נמוכה, דבר המקטין את יעילותם בניסויי הגיבוש. פיתחנו שיטה חדשה העושה שימוש בריכוז אוריאה יחסית גבוה כדי לקבל מסיסות גבוהה יותר של החלבונים. דבר שהוביל אותנו להצלחה בגידול גבישי חלבונים בעלי דיפרקציה איכותית. קיבלנו גבישים מלאים ברזולוציה של 2-2.5Å Data sets שמהם נאספו HspA ו- NblAיציבים של המאפשרת קביעת מבנה תלת מימדי.

התקבלו גם בטיפה תלויה וגם בטיפה יושבת בנוכחות של NblAכל הגבישים של בתור חומר משקע. הריכוז של החלבונים היה בערך 1 מ"ג/מ"ל. 25%ethylene glycol

מהציאנובקטריה NblAאנו מדווחים כאן על המבנה התלת מימדי של החלבון ללא אוריאה 2.5Å ברזולוציה של *Thermosynecococcus vulcanus*טרמופילית של R/R_{free} עד ל Molecular replacementובנוכחותה. קביעת המבנה נעשתה בשיטת 23/24 עבור הגבישים בנוכחות אוריאה ו- 24/29 ללא אוריאה. מהמבנה ניתן לראות כי , מוטיב זה עובר דימריזציה לתוך helix-hairpin-helix בנוי מ- NblA – *Tv*המונומר של אגודה של ארבע הליקסים ביחידה אסימטרית (גבישים עם ובלי אוריאה) בדומה לדימר של *Tv*. במקרה של גבישים מ- Bienert et al שהתפרסם בשנת 2006 ע"י *Anabaena* - NblA עם אוריאה, קיבלנו דימר אחד ביחידה אסימטרית, במקרה של הגבישים ללא NblA – אוריאה, התקבלו שני דימרים ביחידה אסימטרית שביניהם רק אינטראקציה פולארית אחת. ההשוואה בין המבנים עם וללא אוריאה הראתה כי אוריאה לא שינתה את מבנה החלבון, לא גרמה לדנטורציה ואינטראקציות ספציפיות בין החלבונים נשמרו. אנו מאמינים כי שיטה זאת יכולה להיות שימושית מאוד עבור חלבונים קשי לקבלת גבישים בעלי דיפרקציה איכותית.

הראתה הומולוגיה NblAהשוואת רצפים בין חלבונים שונים השייכים למשפחת שונים או של חלבונים ממשפחה זו שמבנם NblAנמוכה. השוואה מבנית של מודלים של כבר נקבע, הראתה כי ישנה הומולוגיה מבנית גבוהה בין החלבונים השונים. לעומת זאת,

השוואה בין הפוטנציאלים האלקטרוסטאטיים של המבנים הראתה חוסר אחידות. ההשוואות הנ"ל מרמזות כי כנראה ישנו קשר של מבנה- תפקוד בחלבונים השונים (structure-related function) .

ופיקוביליזום NblAאנו מציעים מודל תיאורטי אפשרי המתאר אינטראקציות בין יחידת phycocyanin ו- α-helices בנוי משני NblAהשלם. מודל מתבסס על העובדה כי NblA. במודל תיאורטי זה אנו מצעים כי α-helicesהבסיס של פיקוביליזום , עשוי גם כן מ-נכנס לתוך פיקוביליזום בצורה אנכית לעומת אינטראקציה אופקית שהתפרסמה עד עכשיו בספרות.

Synecococus מציאנובקטריה מזופילית NblAהצלחנו לקבל גבישים של *Synecococus elogantus* NblA. הגבישים של 2Å אשר נתנו דיפרקציה של *elogantus* של 48%. partial merohedral twinning עם P4) כנראה בעלי חבורה מרחבית *Se*-NblA) בעזרת molecular replacementניסיונות רבים נעשו על מנת לקבל מבנה סופי בשיטת הפריע לתהליך שיפור המבנה twinningתוכנות שונות. אנו חושבים כי אחוז גבוה של ה-בסביבות R_{factor}). עד עכשיו הצלחנו לקבל רק מודל פרלימינרי של המבנה עם refinement) NblA - ו NblA – *Tv* בנוי מדימרים דומים לאלו של *Se*-NblA40%. המודל הפרלימינרי של *-Anabaena* .

עם סלניום- מתיונין *Se*-NblAלמרות ניסיוננו הרבים לא הצלחנו עד עכשיו לקבל גבישים של MAD .אשר יאפשרו פתירת בעיית הפזות בעזרת שיטת

המומס בעזרת אוריאה נעשו ע"י שיטת הגיבושHspA ניסיונות גיבוש החלבון ובשיטת טיפה תלויה ונסרקו יותר מאלף תנאי גיבוש. גבישים microbatch under oil שונים. גבישים בעלי דיפרקציה PEGהתקבלו בעשרים תנאים שונים, שמרביתם כללו ריכוזי כחומר משקע. 30% PEG MME 550הכי טובה גדלו בשיטת טיפה תלויה עם נאסף מהגבישים הנ"ל. למרות ניסיונותינו הרבים ושימוש 2Å מלא ברזולוציה של data set בעזרת HspAבתוכנות שונות לא הצלחנו לפתור את בעיית הפאזות ולקבל מבנה סופי של לבין יתר HspA. כנראה הבעיה נעוצה בדמיון הנמוך בין molecular replacement שנפתרו עד עכשיו כמו כן קיימת small heat shock proteins החלבונים ממשפחת . עד עכשיו התוצאות הכי טובות HspAאפשרות כי שימוש באוריאה גרם לשינוי במבנה של משיטת

התקבלו ע"י שימוש במודל שנבנה על סמך הומולוגיה molecular replacement השתמשנו רק molecular replacement מחיטה, עבור sHSPמולקולארית עם החלבון באזור

sHSP וה HspA של α-crystalline מהמודל, הדמיון הרצפי בין שני אזורים α-crystalline מחיטה בסביבות 28%. הרבה פרדיקציות של המבנה השניוני נעשו בעזרת תוכנות שונות בדומה β-strains הוא מורכב משבע HspA של חלבון α-crystallineוכולן מראות שהאזור

נותנים את NblA ו HspA מחיטה. קביעת המבנה של חלבוני sHSPלאזור המקביל של ה האפשרות להבנה כיצד חלבונים קטנים מסוגלים להשפיע על ידי הגנה או פירוק של קומפלקסים גדולים כמו הפיקוביליזום.

קביעת מבנה של חלבונים המעורבים ביציבות הפיקוביליזום במקרה של עקה סביבתית

מוניקה דינס

קביעת מבנה של חלבונים המעורבים ביציבות הפיקוביליזום במקרה של עקה סביבתית

חיבור על מחקר
לשם מילוי חלקי של הדרישות לקבלת התואר
דוקטור לפילוסופיה

מוניקה דינס

הוגש לסנט הטכניון – מכון טכנולוגי לישראל
אדר התשס"ז חיפה מרץ 2007

המחקר נעשה בהנחיית פרופ' נעם אדיר
בפקולטה לכימיה

אני מודה לטכניון – מכון טכנולוגי לישראל על התמיכה הכספית הנדיבה בהשתלמותי.