HOCHSCHULE
MITTWEIDA
UNIVERSITY OF
APPLIED SCIENCES

MASTER THESIS

Ms
Maria Beier

Structure development of the heat shock protein 27 by analyzing the evolutionary and functional relationships to other heat shock proteins

2012

Department of **Mathematics, natural and computer science**

MASTER THESIS

Structure development of the heat shock protein 27 by analyzing the evolutionary and functional relationships to other heat shock proteins

Author:
Maria Beier

Course of studies:
Molecularbiology/Bioinformatics

Seminar group:
MO10w1-M

Initial examiner:
Prof. Dr. rer. nat. Dirk Labudde

Second examiner:
Steffen Grunert

Mittweida, August 2012

Did you hear about the rose
that grew from a crack in the concrete?
Proving nature's law is wrong
It learned to walk without having feet.
Funny it seems, but by keeping its dreams,
It learned to breathe fresh air.
Long live the rose that grew from concrete
When no one else ever cared.

The rose that grew from concrete
by 2Pac

Bibliographic description

Beier, Maria: **Structure development of the heat shock protein 27 by analyzing the evolutionary and functional relationships to other heat shock proteins**, 66 pages, 26 figures, Hochschule Mittweida (FH), Department of Mathematics, natural and computer science

Master Thesis, 2012

Abstract

After the expression of the titin-Hsp27-construct with the following purification supplies no satisfied results which makes the realization of the atomic force microscopy not possible. The development of the structure model by using different bioinformatic methods can establish a model for the protein sequence. As bioinformatic methods the template search by different BLAST runs and free available software like SwissModel, Pcons, ModWeb and other tools are used. Nevertheless, the generated model is not the native conformation and has to be analyzed with other software until a stable conformation of the structure can be predicted. Depending on the time which is provided the generated model is a good approach for the aim this master thesis has.

Acknowledgments

First I would like to thank my examiner and adviser Mr. Prof. Dirk Labudde who has teach me the scientific knowledge during my bachelor and master course. He demands the personal assignment of myself and establish the scientific working principle in all of my work.
I would like to thank the Biotechnologisches Innovationszentrum (Biotec) in Dresden and especially the working groups of Prof. Michael Schröder and Ralf Seidel which provide me the laboratory for the expression and purification processes. They also provide the protocols and the plasmids for this scientific work.
I also want to thank my family for always supporting me in my intentions and try everything to solve occurring problems.
Last but not least I want to thank the people who correct these thesis in the development status.

I. Contents

II. List of Figures

III. List of Tables

IV. List of Abbreviations

aa	Amino Acid
ACD	Alpha-crystallin Domain
ATP	Adenosine-5'-triphosphate
BLAST	Basic Local Alignment Search Tool
BVDU	Brivudine
CMT2F	Charcot-Marie-Tooth Disease Type 2F
DNA	Deoxyribonucleic Acid
EMBL	European Bioinformatics Institute
EDTA	Ethylenediaminetetraacetic acid
GO	Gene Ontology
GRC	Genome Reference Consortium
HGP	Human Genome Project
HMN2B	Distal Hereditary Motor Neuropathy Type 2B
Hsp27	Heat Shock Protein 27
IPTG	Isopropyl-β-D-thiogalactopyranoside
MSA	Multiple Sequence Alignment
NAD(P)H	Nicotinamide adenine dinucleotide (phosphate)
NCBI	National Center for Biotechnology Information
PSSM	Position-specific substitution matrix
RNA	Ribonucleic Acid
SSE	Secondary Structure Element
SDS-PAGE	Sodium Dodecyl Sulfate Polyacryamide Gel Electrophoresis
sHSPs	Small Heat Shock Proteins

1 Introduction

The present scientific work has the aim to establish a three dimensional structure of the heat shock protein 27 (Hsp27) and to generate a protein-ligand complex with the therapeutic compound RP101.
Hsp27 is known to protect several cell types against apoptosis when the stress level in the environment is increased. This effect also appears in tumor cells where it obtains to a resistance of the tumor cells against the chemotoxic therapeutics. It achieves that the tumor, which should decrease in size by the chemotherapy, stops this process and starts to grow again. Certainly, in the tumor therapy is such an effect not requested. If an approach could be developed to inhibit the Hsp27 in a way to break the biological function of this protein, the chemo resistance could be probably stopped.
RP101 is an anti-viral therapeutic which is used since the 1970s for the treatment of different herpes virus infections. In different clinical studies an effect against the function of the heat shock protein 27 is established. The resistance against the chemotoxic reactants is determined and the tumor size decreases again. Therefore, no experimental data is generated which gives some information where and how the chemical component RP101 binds to the sequence of the protein. Only some theoretic approaches are developed but without the experimental data is the generated approach only a thesis.
First the aim of this master thesis is the expression and purification of the titin-Hsp27-construct. Afterwards the samples are able to be detached by an atomic force microscopy which generates data of this detachment and associates information about the structural assembly of the proteins. These information will be used for the development of a structure model of the Hsp27 and the complex with the ligand RP101.

Part I

Expression and Purification of a titin-Hsp27-construct

2 Biological Principles

In this chapter I will explain the biological principles of the applied procedures which is necessary to understand why these methods were used. They are listed in their sequential arrangement like they are processed in the laboratory. It is not the aim to generate a protocol for the single techniques in this chapter, only the knowledge behind the processing elements will be described.

2.1 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis is an often used method for protein separation according to their molecular weight and is abbreviated below as SDS-PAGE. Sodium dodecyl sulfate is an anionic detergent which overlaps the intrinsic charge of the proteins. This has the effect that micelles are created which offer a constant negative charge per unit mass. On the gels a potential is applied and the protein samples move from the negative charged cathode to the positive charged anode through the polyacrylamide gel. The substance polyacrylamide is appropriate for the separation of protein samples because it fits to a wide range of protein sizes and can be used for several coloring methods after the separation.

The SDS-PAGE is divided into two single gels - stacking and resolving gel. They have the same ingredients but differ in the concentration and the pH value. For further information the composition of the gels is listed in appendix B. The protein samples are charged with SDS and heat up to 95°C which detaches the tertiary and secondary structure of the proteins by breaking the hydrogen bonds and elongates the whole molecule. With the addition of the reduced thiol compounds, the sulfate bonds between the cysteines are split. In that situation the protein samples create ellipsoids. Figure 2.1 shows the development and the principle of the SDS-PAGE.

2.2 Transformation of C41 Cells

A lot of important compounds for healing are given by nature. Hence, especially for such sciences like pharmacy, it is necessary to produce special molecules that are not natural and afterwards create new drugs. Sometimes it is essential to change the molecules in order to fit better the organism they are generated for or to decrease some side-effects. In all cases a great amount of these products must be processed to sell it as a drug for everyone. This could not always be implemented with methods of the synthetic chemistry so another way should be used. It is possible to create a DNA construct which

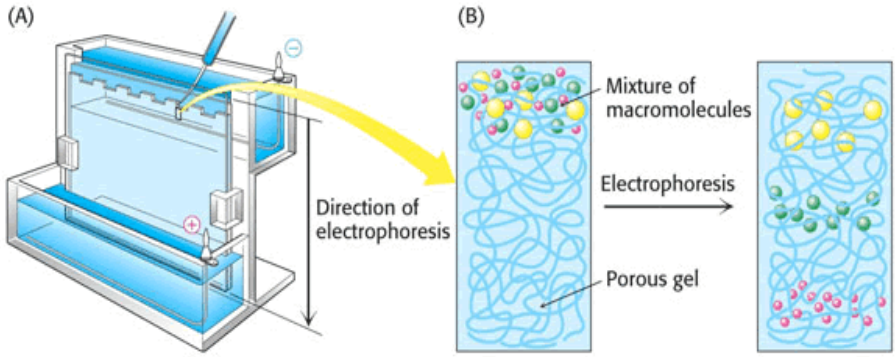

Figure 2.1: The development and the principle of the SDS-PAGE[43]

will be transcribed and translocated into a protein or an other DNA product which, in turn, could be used as the basis for a drug or as drug itself. The construct can be expressed by a cell which has integrated the artificial DNA in its metabolism. The insertion of such foreign DNA is afforded by the method named transformation which is described in this section.

For the project two samples are used. The first one is a control sample, a protein construct containing eight titins labeled as 8xtitin. As protein construct of interest the control sample is changed in a way that the fourth titin is replaced by the heat shock protein 27 in order to be analyzed. This sample is labeled as 7xtitin+Hsp27. To amplificates the pRSET plasmids with the two protein constructs it is necessary to transform them into C41 cells. These C41 cells are mutants of *Escherichia coli* and required for the expression of toxic proteins. If it is planned to express toxic proteins with normal cells it could become difficult depending on the metabolism of the used organism. The worst case is that the toxic protein can damages the metabolism of the host organism and obtains to the programmed cell death called apoptosis of the host. The transformation was realized by electroporation because of it being a timesaving method to make cells competent. The attribute competent means that the cell has the ability to assimilate the foreign DNA into itself. Physical principle of the electroporation is to put cells into an electric field and shot pores into their cell membrane by using short electric pulses of high field intensity. However, it is necessary to work very fast because the pores are closing after a while and no plasmid or other foreign DNA could get into the cells anymore. Figure 2.2 shows the process of electroporation in a schematically way and visualizes the creation of pores by electric pulses.

Often plasmids carry some more information like antibiotica resistance to enhance the chance of integration of the plasmids into the used cells. The cells have the possibility to survive lethal conditions when they carry a resistance against an antibiotica so they assimilate the foreign DNA better. Therefore it is very important to include the needed antibiotica into the culture medium to prevent the plasmids from the throwing-off. This

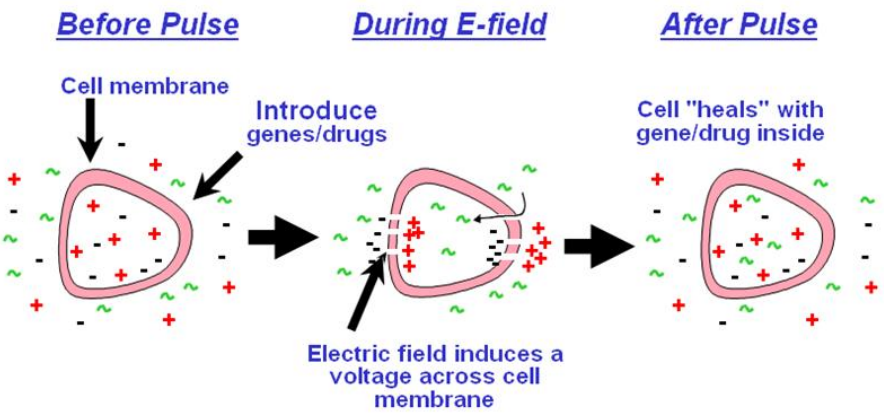

Figure 2.2: The process of electroporation displayed schematically[13]

throw-off of the plasmids can be possible if the antibiotica is not included into the culture medium and the resistance against this antibiotica is not needed anymore.

2.3 Protein Expression

If the foreign DNA or the natural given product is integrated into the metabolism of the host organism the expression process has to be optimized to get the needed products. For the expression of proteins and other DNA products it is advantageous to use a scale-up system. This equates to start with small volumes of expression cultures and enhances it slowly to the volume needed for industrial processing. First a pre-culture is established to guarantee that the cells are adjusted to the culture medium and its conditions. This enabled the cells to adapt themself and their metabolism to the medium before the expression starts. The pre-culture is used to inoculate the culture for the protein expression whose volume is scaled up. The Figure 2.3 visualizes the sequential sub-processes which are done during the protein expression.

The protein constructs within the pRSET plasmids are linked with a promotor which is inactive. It has to be activated before the DNA product could be expressed. In this case the promotor will be activated by Isopropyl-beta-D-1-thiogalacto-pyranoside(IPTG), a molecular biology reagent that controls the transcription of the T7 promotor which is also implemented in the used plasmids. The advantage over promotors which should

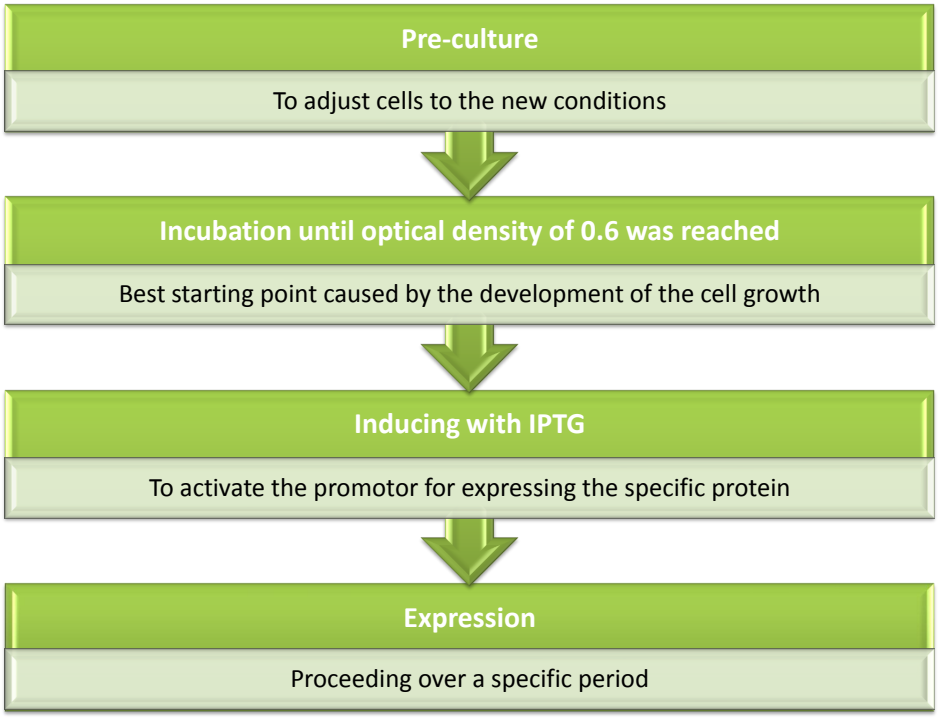

Figure 2.3: Sequential sub-processes used during the protein expression

not be activated is that there is a better control of the expression process. By choosing the best instant of time as starting point for the protein expression good results could be guaranteed. This starting point is detected by the optical density which provides information about the growth curve of the used bacteria. Such curves are generated for static bacteria cultures like *Escherichia coli* and divided into four different stages. The process starts with the so called lag phase where the bacteria adapts itself to the new conditions caused by the medium ingredients and starts to grow. In the following log phase the population doubles every 20 minutes and grows exponentially. This is the optimal starting point for the expression because the cells are more adapted to the medium conditions. Caused by the loss of substrate or the accumulation of the metabolic final products the conditions for growth of bacteria are not good enough so the growth stopped in the third stage called stationary phase. In the fourth period the bacteria dies faster than they could reproduce themselves so their number decrease more and more. The reason is, like in stage three, the metabolic final products which could have toxic effects on the bacteria cells if the amount of end products get out of control. In addition the substrate is not sufficed anymore for the number of cells and the space for growth is utilized. A growth curve of a static bacteria culture with several phases in which the curve could be divided is shown in figure 2.4.

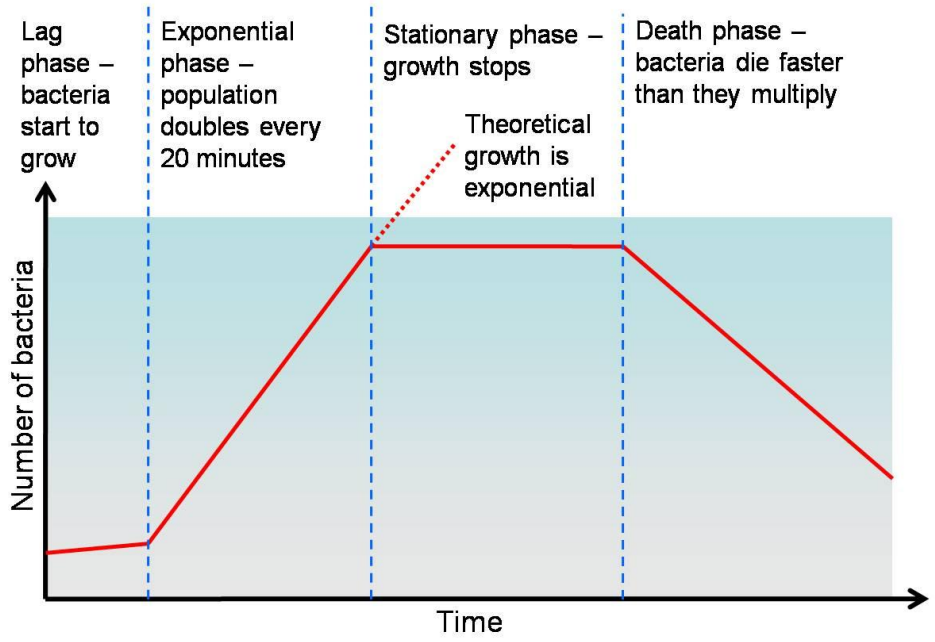

Figure 2.4: Growth curve of a static bacteria culture with the four different stages[17]

2.4 Protein Purification

The protein constructs produced inside of the cell, have to be isolated and purified for further use. First the cell membrane has to be destroyed. Therefore a combination of cell lysis by chemical extraction and ultrasonic is used to guarantee that all proteins inside the cell are extracted. The purification process is done with Ni-NTA beads being little nanoparticles to bind specialized proteins. The proceeding takes place in a column with a matrix where the chemical compounds flow through.

The protein construct is furnished with a his-tag by which it is bound to the beads. They have an anti-his-tag so the protein construct is bound covalently. Hence, the constructs could be isolated from other proteins which are produced inside the cell. To eliminate unspecific proteins which can also be bound to the beads they are washed with a washing buffer. As a last step the bounded protein constructs are eluated by an elution buffer and the flow-through is collected in tubes for further use. In figure 2.5 the different sub-processes of the purification process are displayed.

2.5 Dialysis and Storage of the Samples

In most cases the collected eluates are not used for measurements immediately so they have to be stored until their application. The elution buffer contains different salts

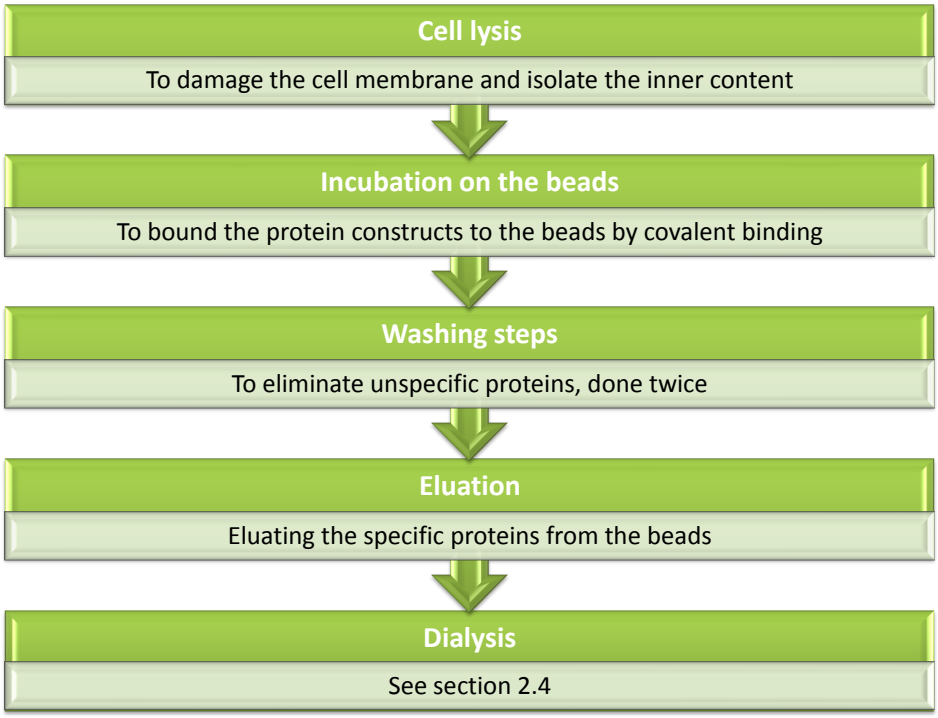

Figure 2.5: Summary of sub-processes contained in the purification process

which could destroy the protein constructs by degrading the bonds inside the protein. Therefore, it is necessary to eliminate the risky compounds by dialyzing the eluates with a special buffer. This buffer is optimized for the protein construct and its conditions. The osmotic gradient is the biological principle which is applied for the dialysis. It is the difference in concentration of two solutions which are separated by a semipermeable membrane. This membrane, composed of cellulose provided with little pores, is able to let small molecules pass through the membrane whereas bigger molecules are unable to transit the membrane. This process is shown in Figure 2.6 to visualize the dialysing process.

Small molecules like the containing salts would pass the membrane and are collected outside the membrane tube. For bigger molecules like the protein constructs of interest the membrane is not diaphanous so they are kept inside the membrane bags. After the dialysis the protein construct could be stored indefinitely without the risk of destroying the protein construct because all salts are passed through the membrane.

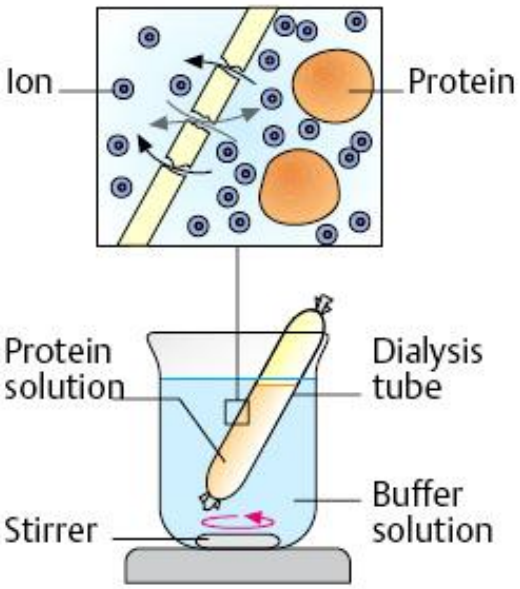

Figure 2.6: Dialysis of the eluates containing the protein samples[12]

3 Methodology

This chapter includes the single protocols for the used methodology which was described in the chapter Biological Principles. General labor equipment is not listed.

3.1 Transformation of C41 Cells

For the Transformation the following material and chemical compounds are used:

- ECM 399 Electroporation System, BTX Harvard Apparatus
- Electroporation Cuvette 0.1cm gap, sterile, Sigma-Aldrich
- C41 cells, 50µl, stored frozen at -80°C
- SOC medium
- petri dishes with LB medium + ampicillin

As prearrangement the frozen C41 cells are thawed slowly on ice while the electroporation cuvettes are kept on ice until they are needed. It is important to pipette 900µl of the SOC medium into 1.5ml tubes and heat them up in a thermo mixer to 37°C before the electroporation. 100ng of each plasmid are added to the C41 cells when they are not frozen anymore. The whole sample volume is pipetted into the cold but dryly electroporation cuvette and put into the ECM 399 Electroporation System. After the pulse is applied, the sample volume is resuspend into SOC medium and put into the thermo mixer. At 600rpm the tubes are rotated for 1hr at 37°C which allows the cells to rest from the stress caused by the electroporation. From each sample 15µl are streaked onto petri dishes with LB medium and ampicillin which are incubated over night at 37°C. It is also possible to create a pre-culture out of the transformed cells like described in section 3.2. The leftover of the transformed cells are admixed with glycerol at the ratio of 1 to 1 and stored in the freezer at -20°C.

3.2 Protein Expression

The following material and chemical compounds are necessary:

- Biophotometer 6131, eppendorf
- Centrifuge Allegra 6R, Beckman
- Incubation Shaker Multitron, inforsakt
- ultracentrifuge Optima LE-80K, Beckman
- 15ml Falcon tubes 2095
- BD Falcon conical tubes 50ml

- 1M ampicillin, aliquot 1.5ml tube
- 1M IPTG, aliquot 1.5ml tube
- 2xYT medium

The protein expression is a two days long process. On the first day a **pre-culture** in 15ml Falcon tubes 2095 is prepared in the afternoon. This is necessary because the bacteria cells should not be incubated for a longer time. A longer incubation time can result in the death of a large amount of bacteria cells and a usage as pre-culture fail. The volume that is used depends on the volume for the expression culture and is applied in a ratio of 1 to 40. Therefore two approaches can be used. First some colonies are picked from a plate by a pipette with a disposable tip and inoculated into 2xYT medium. To make sure that the plasmids with the protein constructs are not dropped out of the cells, the 2xYT medium has to be stowed with 1M ampicillin in a ratio of 1 to 1,000. A second approach for a pre-culture is to inoculate the 2xYT medium with an amount of cells from the glycerolstock in a ratio of 1 to 1,000. The pre-cultures are always incubated over night at 37°C and 220rpm in the Incubation Shaker Multitron.
The second day the **expression culture** is in process. It is started by the inoculation of 2xYT medium with the specific amount of the pre-culture like described above. After the samples are pipetted into 15ml Falcon tubes 2095 they are incubated at 37°C and 220rpm in the Multitron. The period of time for the incubation depends on the volume of expression culture which is used. As a determining factor the optical density is used which is detected by the Biophotometer 6131. Lean on the Lambert-Beer law the optical density is defined as the absorbance of a material in a logarithmic ratio and measured at 600nm. The values must fit in a range of 0.6 to 0.8 which equates to the log phase in the growth curve of bacteria. If this starting point is reached, 60µl are collected as sample for a SDS-PAGE after the expression and 5ml of the expression culture is used as parallel non-induced sample. After the collecting of the samples the expression culture is induced with 1M IPTG in a ratio of 1 to 1,000. Then the expression cultures and the non-induced samples are incubated at 37°C and 220rpm for 3hr. The absorbance is detected a second time after the last incubation to control whether the cells have accreted themself. Like before the incubation with IPTG, 60µl of the expression cultures are collected as samples for the SDS-PAGE. Afterwards the expression cultures are centrifuged in the centrifuge at 3,500rpm and 4°C for 20min to separating the cells from the culture medium. The culture medium is discarded and the cell pellets are stored frozen at -20°C until they are needed for the purification process.

3.3 Protein Purification

For the protein purification process the following material and chemical compounds are used:

- NanoDrop 1000 Spectrophotometer, Thermo Scientific

- Centrifuge Allegra 6R, Beckman
- ultrasonic homogenizer, 2-900ml 70-200W sonoplus HD, Bandelin
- ultracentrifuge Optima LE-80K, Beckman
- centrifuge tubes, Beckman
- Filtration columns, 15ml
- Ni NTA superflow beads, QIAGEN
- Lysis buffer
- Washing buffer
- Elution buffer

The frozen cell pellets are thawed slowly on ice while the lysis solution is compounded. This is necessary because the three different compounds must be mixed fresh due to the enzyme which is used. A complete description of the ingredients is given in appendix B. The cell pellets are resuspend in 10ml of the lysis solution, pipetted into new tubes and incubated on ice for 1hr. Meanwhile the Ni NTA beads are washed because they are stored in 50% ethanol which has detrimental effects on proteins. The required volume of the beads (250µl batch) is washed twice with 5ml of the lysis solution and then put on a rolling table for 10min at 4°C. Afterwards the beads are centrifuged with the centrifuge Allegra at 1,000rpm for 10min at 4°C to separate the lysis solution with the ethanol from the beads. The supernatant is discarded and the beads are stored in a small amount of lysis solution because they had not to dry. After the incubation time of the cell pellets with the lysis buffer the tubes are lysed a second time by a sonicator with 50% of duty cycle until foam come up. The cell homogenate is transferred into centrifuge tubes from Beckman and centrifuged in the ultracentrifuge at 20,0000rpm for 30min. An aliquot of 60µl is taken as a sample for the SDS-PAGE. By pipetting, the supernatant is transferred into the batch and incubated for 1hr by 4°C on the rolling table. After the incubation the batches are centrifuged for 5min at 1,000rpm and 4°C. The emerging supernatant is removed but kept until the purification is finished and no protein can be detected in the eluates. Now the unspecific bounded proteins must be eliminated so the batch is washed twice with 5ml of the washing buffer. The tubes are incubated for 5min on the rolling table by 4°C and then centrifuged at 1,000rpm for 5min. From every supernatant there are 60µl taken as a sample for the SDS-PAGE. After the last washing step 1 to 2ml are left in the batch and the Ni NTA beads are resuspend in it. The suspension is transferred into a column with closed tap and the matrix is sedimented for about 10min. After the incubation the tap is opened and the rest of the washing buffer flows through. Then 2ml of the elution buffer is added to the column with closed tap and incubated for 30min. Afterwards the tap is opened again and the protein is collected in four fractions of 2ml in new tubes. This step is repeated once again and both times an aliquot of 60µl is taken. The concentrations of each fraction is measured with the NanoDrop and could be merged if they are containing protein for the dialysis.

3.4 Dialysis and Storage of the Samples

The following material and chemical compounds are used:

- Dialysis buffer
- Dialysis membrane Spectra/Por 1

The fractions with the protein from the purification process have to be dialyzed to eliminate the different salts like imidazole which could destroy the protein constructs. For this process the dialysis membrane Spectra/Por 1 has to be soaked into the dialysis buffer for over an hour because the membrane is solid at the beginning and will break if it will be modulated. The length of the membrane depends on the volume which should be transferred into it. For 2ml eluate 5cm dialysis membrane are used, every sequentially eluate increase the length about 2cm. When the membrane becomes flexible after the soaking one side is closed by a knot so the other side could be filled with eluates. From each construct the fractions of 2ml are mixed and transferred into a dialysis membrane. The single membranes are closed by a clip and put onto a floating assembly into the dialysis buffer. The beaker glass with the dialysis buffer is put on a stirring device at under 300rpm and dialyzes overnight.

4 Results

The results of the single working methods in this chapter and their meaning for the whole project of the expression and purification of the titin-hsp27-construct will be presented.

4.1 Transformation of C41 Cells

First the transformation is done with about 40ng of the two different plasmids because this option is used in the protocols before, supplied by the Seidel group of the Biotec in Dresden. However, the expected clear results in the SDS-PAGEs could not be proven. During different gel electrophoresis with the plasmid material no errors occur making the transformation itself likely to be the problem. Hence, the transformation is done with 100ng of each plasmid which generates the expected results. Figure 4.1 shows the gel electrophoresis results of the two plasmids digested by the restriction enzyme AlwNI. The plasmid with the 8xtitin construct has only one binding site for the restriction enzyme AlwNI resulting in a 5000bp fragment like detected in the following figure. In contrast the plasmid with the 7xtitin+Hsp27 has three binding sites so three fragments occur with sizes of 437bp, 1734bp and 3247bp. The non-digested samples show the typical three bands for non-digested plasmids which occur in the following succession as relaxed, linear and coiled plasmid.

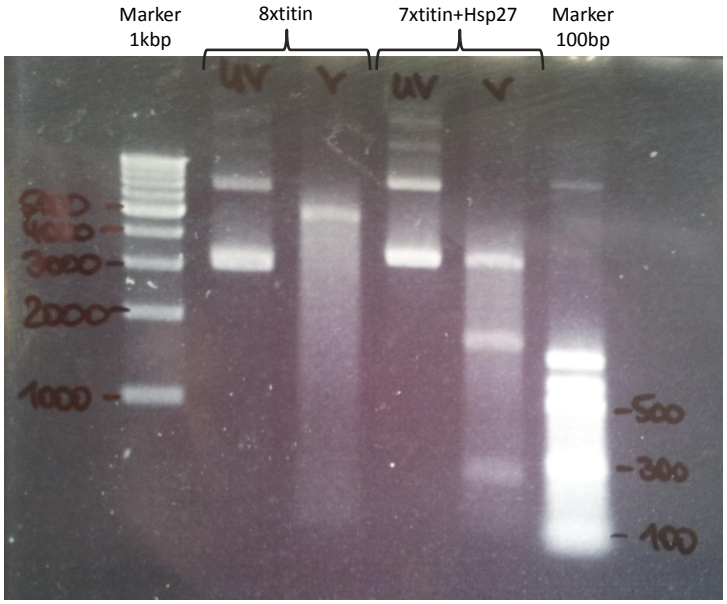

Figure 4.1: The gel electrophoresis of the two plasmids digested by the restriction enzyme AlwNI

4.2 Protein Expression

For the first time the SDS-PAGEs for testing the expression process show no clear expression bands for both plasmids. These expression bands should appear at 80kDa for 8xtitin and at 97kDa for 7xtitin+Hsp27. After the enhancement of the plasmid volume, used for the transformation, brighter expression bands at the 8xtitin samples appear. The parallel incubated and non-induced samples indicate the same bands and authenticate the results that way. For the expression of the 7xtitin+Hsp27 no bands could be detected over the whole period. As a possibility to optimize the expression of the protein constructs, the amount of IPTG is doubled so that the new ratio is 1 to 500. Unfortunately on the one hand the expression bands in the SDS-PAGEs become more clear but on the other hand the right expression bands for the 7xtitin+Hsp27 could not be proven. A sequence analysis has detect some gaps in the sequence of the Hsp27 which could be the reason why no expression process is started. In figure 4.2 the results of the SDS-PAGE of the expression samples colored by Coomasie are displayed. As blank value the samples before the inducing of the IPTG are used.

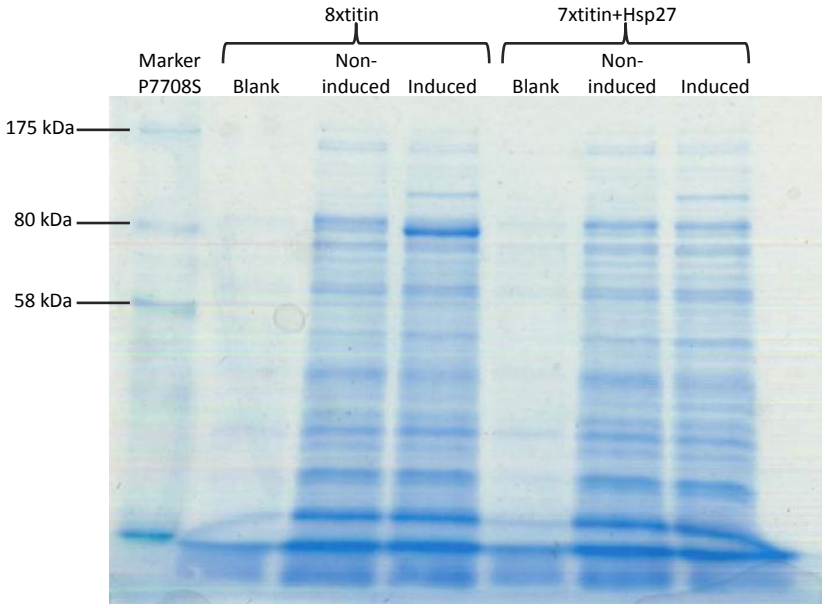

Figure 4.2: SDS-PAGE of the protein expression samples colored by Coomasie

4.3 Protein Purification

After the appearance of the brighter expression bands in the SDS-PAGEs at the 8xtitin, some purification processes are started even if no bands could be detected at the 7xtitin+Hsp27. However, no 7xtitin+Hsp27 could be assured by SDS-PAGEs in the eluates after several purifications. As a result of the loss of the important protein construct

the washing buffer is changed in its chemical composition. This buffer is used to eliminate the unspecific bound proteins. It could be possible that also the protein construct is detached from the beads. Different modifications of the washing buffer show no essential improvement so it could be assumed that already the expression of the protein failed. Figure 4.3 shows the SDS-PAGE of the purification samples from the 8xtitin colored by Coomasie where clear bands by the size about 80kDa could be detected so the 8xtitin is expressed and purified.

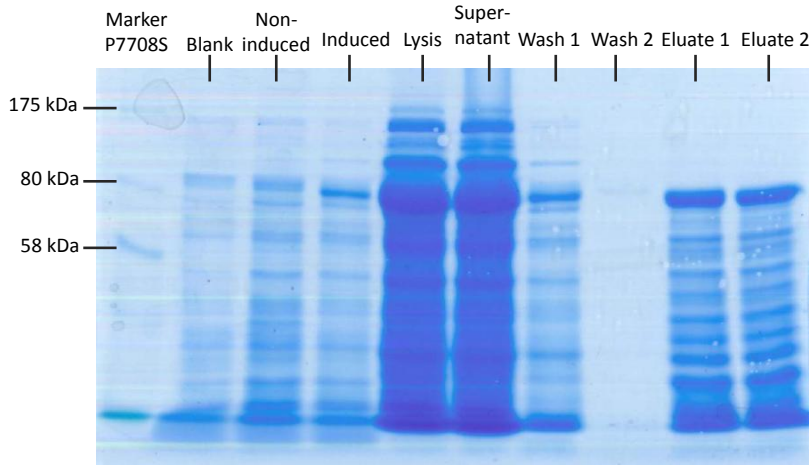

Figure 4.3: Protein purification samples in a SDS-PAGE colored by Coomasie

Figure 4.4 shows in contrast the SDS-PAGE of the purification samples of the 7xtitin+Hsp27, where no bands of a size about 97kDa are detected in the eluates so the expression probably failed. Only some fragments could be detected but it is not clear if these fragments contain the 7xtitn or the Hsp27. In the supernatant and the lysis samples of the 7xtitin no bright bands about 97kDa are detected.

4.4 Dialysis and Storage of the Samples

The first dialysis is started with water as a dialysis buffer because no appropriate buffer could be identified. Hence, the different salts like imidazole in the buffers especially in the elution buffer could destroy the protein constructs so water should not be used as dialysis buffer. As good dialysis buffer a combination of PBS with 0.1mM EDTA could be established even if there could be detected some fragments in the eluates. This is possible because the protein constructs are attached by an his-tag at the beginning of the constructs so incomplete fragments could also be bound.

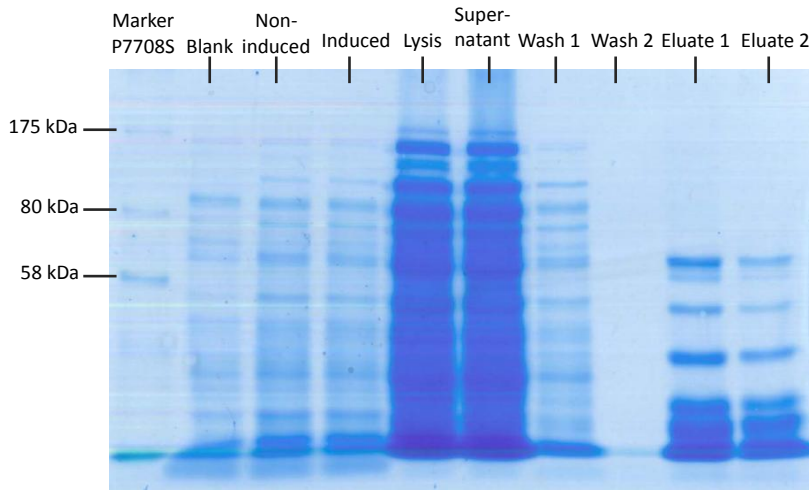

Figure 4.4: SDS-PAGE of the protein purification samples of the 7xtitin+Hsp27

5 Discussion

In this chapter the results will be discussed and some answers will be given for problems that occur during the project with the titin-hsp27-construct.
The transformation could probably also generate satisfying results with a lower concentration of the used plasmids. Hence, the option with 100ng gets a better production because there is more foreign DNA material which could be assumed from the cells. Also the SDS-PAGE gels fleshed out this thesis so 100ng could be detected as adequate amount for the transformation. The expression could only be completed for the 8xtitin not for the 7xtitin+Hsp27. Possibly reasons are the gaps in the sequence detected by the sequencing analysis. If the amount of gaps is too high, the polymerase could not bind to the DNA and start the transcription. Hence, no protein construct could be translated because no mRNA is usable as template. Another possibly reason can be that *Escherichia coli* is not an appropriate host organism for the expression. Maybe the protein construct is too big and another host organism like yeast is better for the expression. A lot of publications use the expression in mammalian cell lines which offers probably one of the most difficult cultivation conditions. Because of the short period of time no other hosts are tested. Also the cloning of a shorter construct with less titins could not be tried by the short spell but could become a possibility.
Given the fact no protein construct could be expressed and purified it is not possible to realize the atomic force microscope measurements so the structure of the heat shock protein 27 must be developed by theoretical considerations by bioinformatic methods described in the following part II.

Part II

Structure Development by Bioinformatic Methods

6 Sequential Analysis of the heat shock protein 27

This chapter gives attention to the sequential analysis of the heat shock protein 27 and contains all information which could be generated by the sequence as a basis.

6.1 General Information about the gene HSPB1

Heat shock proteins in general are first established in 1962[42] as a set of highly conserved proteins whose expression is induced by stress[15]. Most of the heat shock proteins have a strong cytoprotective effect and are ATP-dependent chaperones[15]. Especially the small heat shock proteins (sHsps) act in a ATP-independent variation[15] The heat shock protein 27 belongs to the family of sHsp which are all characterized by the low molecular weight between 12 and 43kDa[6]. They protect the cell against apoptosis and appear in all mammalian cell types. sHsps act as molecular chaperones which antagonizes the formation of aberrantly folded proteins and allow their correct refolding[6]. All Hsps are expressed on a common level in the cells and are involved into several cellular processes. Some of these processes are intracellular transport, cytoskeletal architecture, translation regulation and intracellular redox homeostasis[6]. Ten different small heat shock proteins are characterized in human but only Hsp27, Hsp22 and alphaB-crystallin are true heat shock proteins which are expressed on a higher level when the cell stands under stress conditions or other environmental changes[6]. The small heat shock proteins own a conserved domain, a so called alpha-crystallin domain (ACD). They also contain a N-terminal WDPF domain[6]. After the conserved ACD a non conserved flexible domain appears in the C-terminal region of the protein. The sHsps have the property to form globular oligomeric structures which could reach molecular masses ranging from 50 to 700-800kDa[6]. It is assumed that these oligomers control the activity of the proteins. All sHsps have the ability in response phosphorylated which modulates their activity in response to several stimuli[28, 47]. Specially Hsp27 has three phosphorylated serine sites, one is located in the WDPF domain at position 15[47]. The two other phosphorylation sites are associated with the serines at the positions 78 and 82[6]. The following figure 6.1 shows the domain architecture of the Hsp27.

For Hsp27 could be demonstrated that the phosphorylation results in a decreased size of the oligomers[22]. These oligomers are responsible for the tumorigenic activity of Hsp27[8].

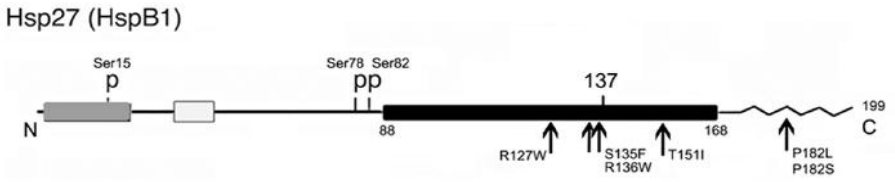

Figure 6.1: The domain architecture of the Hsp27[6]

6.2 Analysis with the Basic Local Alignment Search Tool

6.2.1 Basic Local Alignment Search Tool

The Basic Local Alignment Search Tool is a searching tool based on the sequence homology, identified by the dynamic programming Smith-Waterman algorithm[5]. A number of software tools, developed before 1990, use the measure of similarity of sequences to detect biological relationships between two sequences[5]. Therefore, a better result could be reached by using the dynamic programming algorithm developed by Needleman and Wunsch[33] or Smith and Waterman[44]. With the assistance of such algorithms it is possible to search large databases for sequence homology without special purpose hardware[5] because the request of computing time and storage space is not too high for common computers.

BLAST starts to divide the query sequence into substrings of the length w which are validated by a score T. For protein sequences the length is in general 3, for nucleotide sequences is mostly 11 used[31, 18]. By comparison the query sequence with the substrings are all words, following called w-mers, with the length w detected so sufficient similarity could be identified[5]. Then the algorithm starts for every sequence B of the database and search the established w-mers[31]. For every hit the position would be determined. This accords to the first step of the BLAST approach to localize the hits or segment pairs like the short subsequences are called[18]. Within the second step it would be identified which pairs of hits lay on a diagonal line of a thought matrix and have a distance smaller then L (for proteins preset to 40)[18]. The distance is calculated as difference of the position of each first character of the w-mers and describes the window size L[31]. Only if there is a second hit within this window size the two hits are recognized in the following course[18]. After a second hit is found both will be expanded bidirectional until the score cannot be increased anymore[18]. If the score reaches a threshold S, an expansion with gaps will be preferred[31]. The resulting alignment with gaps would be used as output if the calculated E-value is lower then a determined threshold[31]. This E-value is a parameter which describes the number of hits the user can expect to see by chance and is decreased exponentially by the increasing of the

bit score for the BLAST alignments[30]. Even lower the E-value is, the more significant the match in the BLAST search is. The bit score gives an indication how good the alignment of two sequences is[30]. Even higher the score is, the better the alignment is. It is calculated from a formula which takes account of the alignment of identical and similar residues, as well as the inserted gaps[30]. In figure 6.2 the algorithm of the Basic Local Alignment Search Tool is visualized.

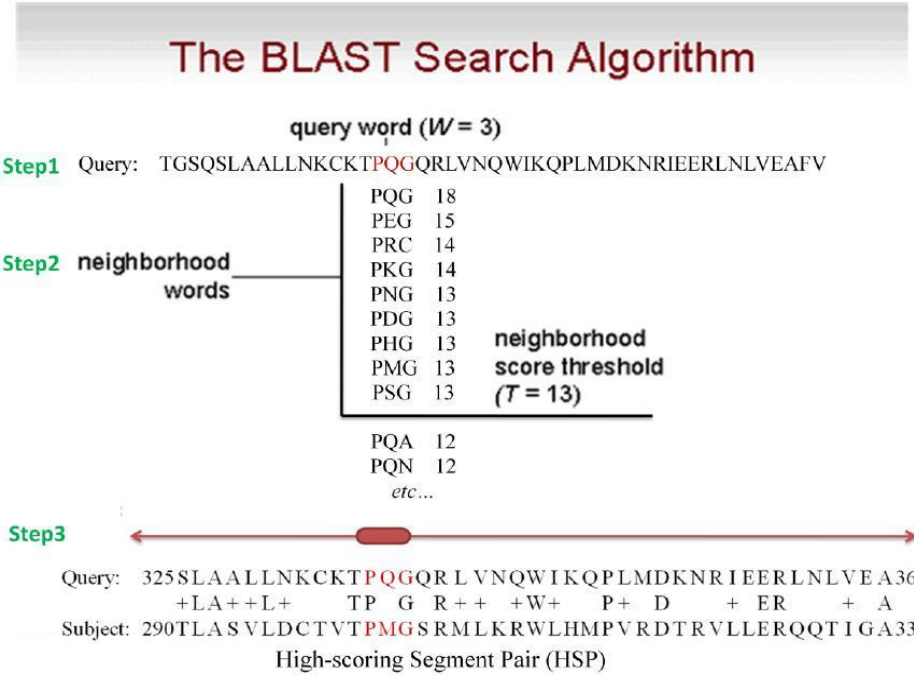

Figure 6.2: Visualization of the BLAST algorithm[7]

6.2.2 Results of the BLAST approach and the corresponding properties of the Hsp27

By using the BLASTn the following information could be collected by analyzing the generated hits of the blastn run. The user has different variations he can chose from to be used for the search tool. A megablast run establish hits by using only high similar sequences whereas the normal blastn run search in somewhat similar sequences and is able to identify more hits but with a lower sequence identity[30].
The heat shock protein 27 is coded by the gene HSPB1 which is deposited in the GeneBank under the gene ID 3315. In turn the gene is located on the chromosome 7 of *homo sapiens* whose reference is GRCh37.p5. GRC is the abbreviation for Genome Reference Consortium whose assignment is to put the determined sequences in the right chromosome context to describe the allelic diversity of the sequences. It is a cooperation of the National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EMBL), the Genome Institute at Washington University and the

Wellcome Trust Sanger Institute. GRCh37 is the human reference genome released in May 2009 in cooperation with the human genome project (HGP). p5 is the abbreviation for patch release 5 in which patch stands for a scaffold sequence being part of a genome release, in this case the release number 5. The sequences correct errors in the chromosomal assembly or add additional loci and show separated regions of the chromosomes.
As concrete location for the gene HSPB1 is known as 7q11.23 which describes the locus on the chromosome 7, stored in the q tail under the id 11.23. Figure 6.3 displays the region of the HSPB1 on the chromosome 7 and the corresponding transcription products.

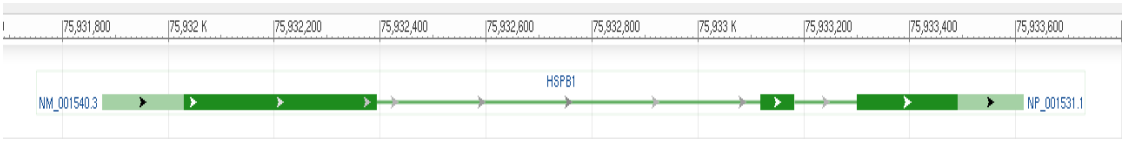

Figure 6.3: The region of the gene HSPB1 on the chromosome 7 and its transcription products deposit by the NCBI

Table 6.1 gives a summary of the different IDs that belong to the heat shock protein 27 or the corresponding gene HSPB1. The NCBI collects all information about the gene HSPB1 and the chromosome it is located on as well as the resulting DNA products.

Database	**ID of Hsp27 or HSPB1**
GeneBank	3315
UniProt	P04792
SwissProt	P04792
NetPath	M3315
NCBI	Chromosome: NC_000007.13
	Gene: NG_008995.1
	mRNA: NM_001540.3
	Protein: NP_001531.1
MIM	602195
HGNC	5246
HPRD	09076

Table 6.1: Summary of the IDs belonging to Hsp27 or HSPB1

HSPB1 has a complete length of 1,740bp and is transcribed in 914bp mRNA whose translation leads to the 205 amino acids (aa) long protein Hsp27.

With the help of BLASTp it is possible to find sequence homology and identify similar sequence regions with almost similar function. In the case of unknown tertiary structures

of a new sequence the homology could be used to establish some structure relationships by searching a protein database. Hence, it is possible to discover functional relations, sequences and from this structure similarity and finally evolutionary relationships. PSI-BLAST is a special version of the blastp where a position-specific iteration BLAST algorithm is able to find iterative protein sequence similarities[30]. Based on a normal BLAST run a multiple sequence alignment is calculated and a consensus sequence is established[18]. Therefore, a position-specific substitutions matrix (PSSM) is generated with Lx20 fields[18]. This breed a targeted search for related protein sequences because the conserved region obtain a higher score while non-conserved regions achieve lower scores[18].

By a BLASTp run one structural and functional domain of the Hsp27 is identified. This domain is called the alpha-crystallin domain, abbreviated by ACD, which is a polypeptide binding site. The N-terminal region before and the C-terminal region after the alpha-crystallin domain is very flexible in sequence and length[6]. The appearance of this functional domain results in an assignment into the Hsp20 or alpha-crystallin family which is deposited under the ID pfam00011 or 200926. Their members are counted to the superfamily of the alpha-crystallin p23 like heat shock proteins. This superfamily is based on a domain which is similar to the alpha-crystallin domain of the small heat shock proteins. However, it belongs to the p23, a cochaperone for the heat shock protein 90. It shows the special relation between the small heat shock proteins like Hsp27 and the bigger heat shock protein 90.
The first best fitting hits of the BLAST run are the sequences which have the highest similarity to the query sequence depending on the algorithm used by BLAST. As the best fitting sequence with a similarity of 91% was 3Q9P detected, a HSPB1 fragment which contains the alpha-crystallin domain with a gap. Unfortunately, it does only cover a region of 85 amino acids in the range of aa 90 to 171. The following hits in table 6.2 did not reach the high similarity of 3Q9P but fit to other regions.

PDB ID	Query Coverage	Score	E-value	Sequence Identity
2KLR	83%	57.4	1e-10	45%
3N3E	43%	49.3	4e-08	54%
3L1E	47%	48.5	7e-08	54%
3L1F	51%	43.9	4e-06	50%
2Y1Z	49%	42.7	8e-06	50%
2WJ7	43%	42.7	8e-06	53%
2Y22	43%	41.6	2e-05	52%
3L1G	39%	39.3	1e-04	56%
2Y1Y	48%	38.5	2e-04	50%
2WJ5	35%	30.4	0.27	56%

Table 6.2: Summary of the BLAST hits with the highest score and sequence identity

Most of the established hits show a sequence identity of about 50%. Therefore, the hit

with the highest range is the structure of 2KLR which covers 83% of the query sequence. Some structures with a better sequence identity are 3L1G and 2WJ5 with 56%. The Bit score and the E-value of these hits are therefore really low and the range which is covered of the query sequence reach only 39% and 35%. This approved the statement that the sequences of the Hsp27 and the heat shock proteins in general vary a lot in sequence with the exception of the conserved alpha-crystallin domain. Such facts could be a problem for the modeling process.

By using the PSI-BLAST, there are generated more related protein structures. Such relationships are established by conserved regions in the sequence. Hence, evolutionary successful sequence motifs are stored as these conservations. They are often associated with a special function or a specific three dimensional structure which is kept by organisms for several generations. At the current point of time the sciences assume that the function of a protein is not mandatory associated with a specific folding motif[26, 56]. This thesis is proven by different aspects. A large number of enzymes belong to the folding class of the TIM-barrel but own different functions. Hence. the folding motif of the TIM-barrel is effective so it is stored as conservation but it does not carry the function[26]. All these enzymatic structures have a common ancestor and then developed divergent[26, 18]. Nevertheless, the function of a protein is also associated by a conservation and therefore by a special motif[56]. Hence, the active center of enzymes belonging to the same enzyme class is located at the same position in the three dimensional structure[26, 56].
Therefore, the motif search can establish important information about the protein function or the folding structure. In the case of Hsp27 four different sequence motifs could be established. They are listed in the following table 6.3.

Motif ID	**Description**	**Containing IDs**
IPR001436	Alpha crystallin/heat shock proteins	PR00299, PIRSF036514
IPR002068	Heat shock protein 20	PF00011, PS01031
IPR008978	Hsp20-like chaperone	SSF49764
noIPR	unintegrated	G3DSA:2.60.40.790, PTHR11527(:SF33)

Table 6.3: Sequence motifs generated by InterProScan

The motif IPR001436 contains different motifs of the crystallin and the heat shock protein 20, established by several databases like PRINTS, ProSite, Pfam and InterPro. Crystallin is a water soluble protein which occurs in high concentration in the cytoplams of eye lens fiber cells[46]. This functional group is divided into four subgroups called alpha-, beta-, gamma- and delta-crystallins. The classification is based on the charge, the size and the immunological properties[46]. The first motif is stored under the name Acrystallin which is a 7-element fingerprint and provide a signature for the alpha-crystallins[46]. It is generated of a alignment containing 10 sequences and drawn from the conserved regions over the whole alignment length[46]. Therefore, the motif

covers a region of 169 amino acids established with several gaps. These gaps can appear by the fact that the sequence is a consensus sequence established by different query sequences. In the consensus sequence are the most probably residues collected which are generated by the frequency of occurrence in the query sequences. The second ID deposit the Hsp27 and cover the query sequence from amino acid 6 to 205.
By the ID IPR002068 are two motifs collected which belong to the heat shock protein 20. It contains the family of the heat shock protein 20 like also identified before. The superfamily is deposit by the PDB ID 1SHS, a small heat shock protein from *Methanococcus jannaschii*. Confirmed by different sources is the role of heat shock proteins in organisms as response to heat shock and other environmental stress. All family members have an average molecular weight of 20kDa and act as molecular chaperones to protect other proteins before apoptosis[38].
A superfamily called Hsp20-like chaperone is established with the ID IPR008978. It contains different GO terms which are described in the section "Functional Placement of Hsp27". The enzymatic classification by EC terms shows that the Hsp20 family act on NADH and NADPH. They also need a heme protein as acceptor, but continuative information could not be established. There could be some relations annotated to some limitations of the human physical health like muscle weakness and different abnormalities of muscle physiology and morphology. The motif covers a region of the alpha-crystallin domain from amino acid 74 to 182.
The last motif has no IPR and contains three different IDs. The ID G3DSA:2.60.40.270 belongs to an item of the Gene3D database and is connected with the CATH database. It categorizes the motif to a special structure class deposit by the first three numbers within the ID. The number 2 stands for "mainly beta" and describes the folding class validated by the established secondary structure elements. Therefore, the main part of the query structure contains beta-strands. By number 60 a architecture motif called sandwich is lodged. Hence, it is a combination of two secondary structure elements with different properties like helix and beta-strands in the domain architecture. The third number, in this case 40, describes the topology of the protein structure and is deposit by immunoglobuline-like which is found in many proteins. Immunoglobulin-like proteins contain beta-strands which are located in a twisted way[9]. This is finally called beta-sandwich. Hsp27 is therefore a immunoglobulin-like protein with a beta-sandwich as established domain architecture. The results of the modeling need to be awaited to answer the question whether this domain architecture appears. The motif covers the query sequence from amino acid 85 to 183. By the IDs PTHR11527 and PTHR11527:SF33 deposit two items of the PANTHER database which contains the Hsp20 family and especially the heat shock protein 27. This motif reach the whole query sequence from amino acid 1 to 205.
The following figure 6.4 shows the established motifs by InterProScan with the belonging IDs.

There are detected more structures which are probably related to the Hsp27. Table 6.4 shows the five structures with the covered region to the query sequence.

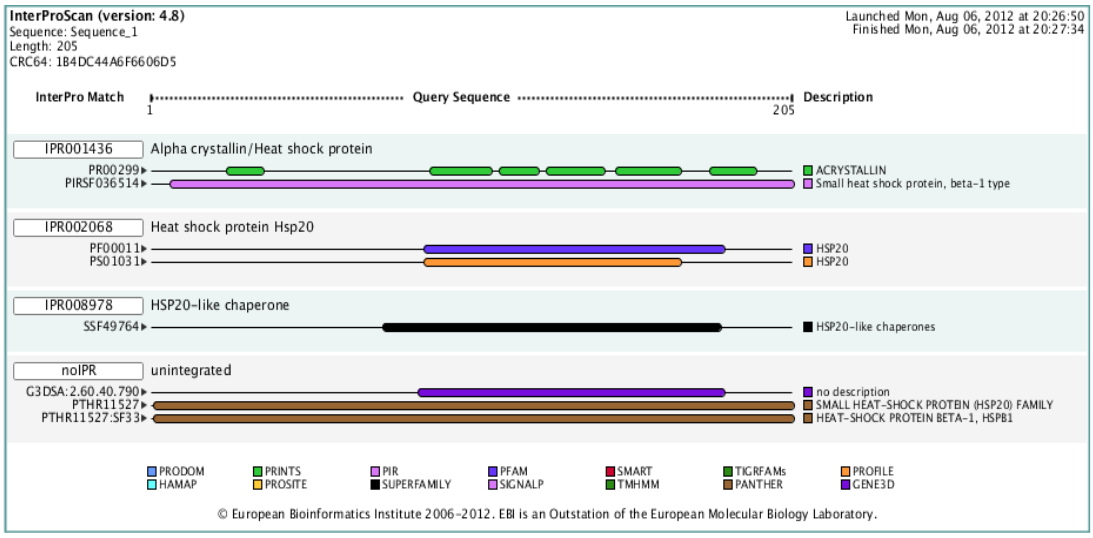

Figure 6.4: Conserved motifs established by InterProScan

PDB ID	Chains	Alignment Length	Sequence Identity	E-value
2KLR	A,B	171aa	45%	2.11e-34
2YGD	A-X	171aa	45%	2.11e-34
3N3E	A	105aa	50%	2.25e-28
4ELD	A	54aa	41%	5.24e-03
2BOL	A	76aa	28%	0.2

Table 6.4: Related structures to the query sequence of Hsp27

On the basis of the E-value these five structures show a relation to the heat shock protein 27. 2KLR is a solid state NMR structure of the alpha-crystallin domain in some alphaB-crystallin oligomers and 2YGD is the structure of the 24meric eye lens chaperone alphaB-crystallin. These alphaB-crystallin is also a small heat shock protein and is expressed during heat shock. They cover the widest range of the query sequence but have only an identity of 45%. Also 3N3E, the alpha-crystallin of the zebrafish, and 2BOL, the crystal structure of the metazoan small heat shock protein Tsp36, are detected by the BLAST search which was done before. The Tsp36 is a tapeworm sHsp with a duplicated alpha-crystallin domain[21]. Only 4ELD is a new hit and deposit the crystal structure of an activated variant of the Hsp16.5. For 4ELD a lower E-value was calculated so it is not established as a hit in the Basic Local Alignment Search Tool. Hence, the sequence identity of 41% over a range of 54aa is able to deliver new information for structure modeling.

A multiple sequence alignment of the generated related structures with the query sequence of Hsp27 shows a region of higher sequence identity. The alignment contains a lot of gaps but the conservation of the alpha-crystallin domain appears clearly. Before and after this region of high conservation no other conserved clusters are detected. So it seems that the modeling process of the structure will become difficult, caused by the flexible areas around the conserved domain. Figure 6.5 shows these conserved

regions.

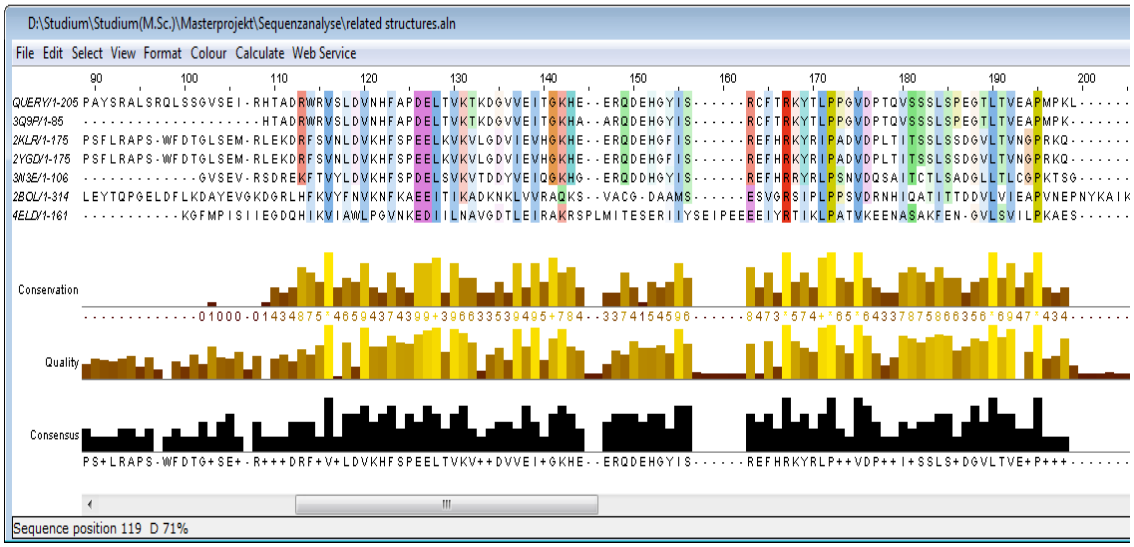

Figure 6.5: The conserved region of the related structures established and colored by ClustalX

The database UniProt has some information about the secondary structure of the heat shock protein 27. Hence, the information are only for the region of the alpha-crystallin domain. A summary of the detected secondary structure elements with the corresponding amino acid region is given by the following table 6.5.

Secondary Structure Element	Amino Acid Region
beta-strand	94-99
turn	102-105
beta-strand	110-113
beta-strand	118-121
beta-strand	140-142
helix	150-152
beta-strand	154-157
beta-strand	163-168

Table 6.5: Secondary Structure Elements identified within the AC domain

The number of beta-strands points to a beta-barrel as folding motif. This could authenticate the architecture beta-sandwich deposit in the CATH database. Conspicuous is the short range of each listed secondary structure element which seems to be not correct.

6.3 Phylogenetic Analysis of Hsp27 and all heat shock proteins with Phylip

Based on the results of the BLAST approach, the different Hsp27 genes are analyzed about their phylogenetic information. The alignment of two sequences show similar as well as identical sequence regions and generates a distance between the sequences which should express their similarity. If these sequences are supposed to be homologue the distance could be seen as an evolutionary distance and gives information about the evolution of the genes itself[20].

6.3.1 Phylip

Phylip is the abbreviation of PHYLogeny Inference Package and collects a program package for evolutionary trees[39]. It is free available and written to work on many different computer systems[39]. The packages are written in C and could be changed by the user[39]. There are also executable packages for Windows, MAC OS X and Linux systems receivable which is a good option for users without the necessary programing knowledge[39].
Each program is controlled by a menu where the user can change the listed parameters[39]. As input data can be used molecular sequences, gene frequencies, restriction sites and fragments, distance matrices and discrete characters[39]. The used programs are protdist, protpars, consense, drawgram, drawtree and retree. They are also listed in their sequential course and the table 6.6 describes the single method which belongs to the programs.

Program	Belonging Method	Output
protdist	calculates distances between protein sequences	output file
protpars	generates different phylogenetic trees by parsimony method	output and outtree file
consense	generates consensus tree out of the trees generated by protpars	output and outtree file
drawgram	draws rooted trees out of an outtree file	user-defined
drawtree	draws unrooted trees out of outtree file	user-defined
retree	reads outtree files and allows modifications of the generated trees including the change between unrooted and rooted trees	outtree file

Table 6.6: The single programs with their described method

Unfortunately, it is not possible to deliver the calculated distance matrix to the protpars program as input. This results in phylogenetic trees without distances on the differ-

ent branches which makes the interpretation of the generated trees a little bit difficult. Therefore, the phylogenetic trees are generated additionally by Jalview with a neighbor joining method and a BLOSUM62 matrix. Given by the fact that the trees generated by Phylip use also the neighbor joining method they are equals to the Jalview trees.
The phylogenetic analysis are done with the Hsp27 and also with all related sHsps. The results are described in the following sections.

6.3.2 Phylogenetic Analysis of Hsp27

Figure 6.6 show the generated phylogenetic tree. The following results are based on a PSI-BLAST run of the query sequence of Hsp27. By changing the E-value from general 10 to a smaller threshold like 1, 0.1 and 0.01 the generated results can be filtered. All established hits can be regarded as significant hits because the probability that a hit is found randomly is equals to the determined E-value. For this phylogenetic analysis the results generated with the E-value 1 are used because all hits are also established by the other E-values and do not differ at all.

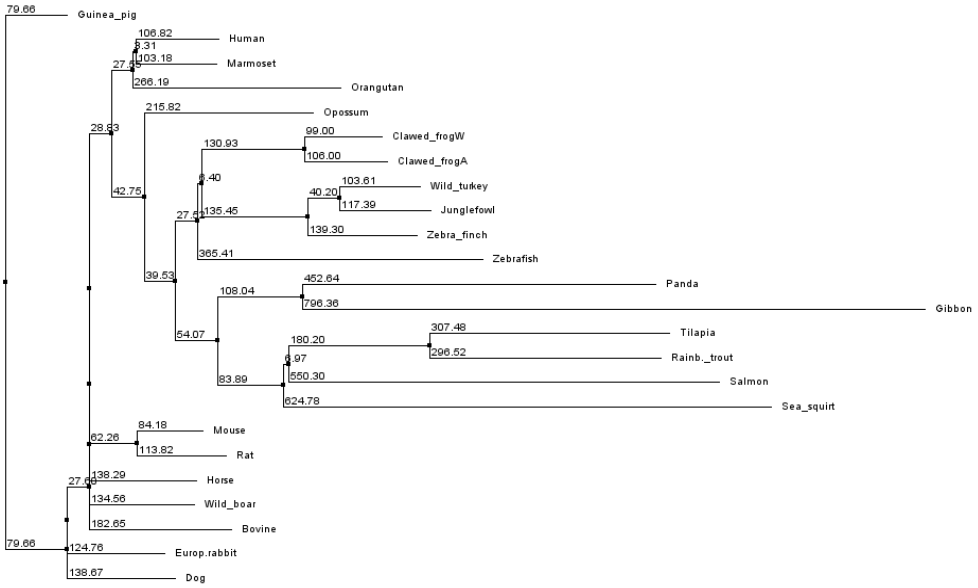

Figure 6.6: Phylogenetic tree of the different Hsp27 sequences generated by Jalview

The human heat shock protein 27 is located in a little group of two sequences in the upper region of the phylogenetic tree. This group contains among the human sequence the sequence of the common marmoset. However, the sequence of the orangutan is placed in a leave related to the leave which is split into the group containing the Hsp27. Interesting is the fact that the sequences of the human Hsp27 and belonging protein of the common marmoset are evolutionary closer and the sequence of the Hsp27 of the orangutan offers the greater evolutionary distance. By using the binary taxonomy based on Carl von Linné the sequences of orangutan and human are related closely

then the sequence of the marmoset. Orangutan and human belong to the same family, the huminidae. The common marmoset therefore belongs to the family of the callithrichidae and is distantly related to the other two sequences. This obtains the conclusion that the sequences of the protein are more related then the involved organisms.
The previous leave, which contains the group of the primates, has a second branch which obtains to big group with several subgroups. These subgroups contain amongst other things the group of the birds and the frogs. The group of the birds contains the sequences of the red junglefowl and the wild turkey in a subgroup. These two birds belong both to the family of the phasianidae which declares the evolutionary close relationships of the heat shock protein 27 sequence. Therefore, the third sequence of the bird group belongs to the family of the estrildidae and is a little bit distantly related to the other two sequences. The upper group of the frogs has its source in the same leave which is the source for the group of the birds. In the frog group are only the sequences of the western clawed frog and the african clawed frog. Both sequences are very close relative depending on the fact that they belong to the same genus. The source leave of these two groups is related to the sequence of the zebrafish. This is the only fish which is not integrated into the group of fishes in the lower region of the phylogenetic tree. Reasonable for this exclusion is the fact that these sequences share only the same class and no closer relation. The previous source leave contains two groups - the already respond group of the fishes and a smaller group with the gibbon and the panda sequences. The last group has the same class, the mammalian. An interesting question is, why the gibbon sequence is not integrated into the group of the primates in the upper area of the tree. The gibbon also belongs to the primates. Given by the fact that this phylogenetic tree is calculated in the sequence similarity of the alignment which is generated by BLAST. The deposit sequence is also only predicted but the other primate sequences are also only predicted. Hence, maybe the sequence own some gaps more then the other primates sequences which results in the exclusion from these group. The group of the fishes contain four sequences. These sequences belong to the class actinopterygii and embrace the organism of the atlantic salmon, the nile tilapia, the rainbow trout and the seq squirt. The last one is separated from the other three sequences because the only similarity is the phylum chordata.
The group of the mammalian species is located in the lowest region of the tree and has only a branch to the leave which is the source for the primates group and the other great group which is described above. All of the containing sequences are distantly related which suggest that these sequences are very similar. And in fact the sequences show a great sequence identity, especially in the region of the alpha-crystallin domain.

6.3.3 Phylogenetic Analysis of the related sHsps

Figure 6.7 shows the phylogenetic tree for the small heat shock proteins, generated by Jalview. The results are based on a PSI-BLAST run with an E-value of 0.1 because a higher E-value of 1 established too much hits. The Hsp27 sequences from the phylo-

genetic analysis described above are excluded because these hits are then repeated.

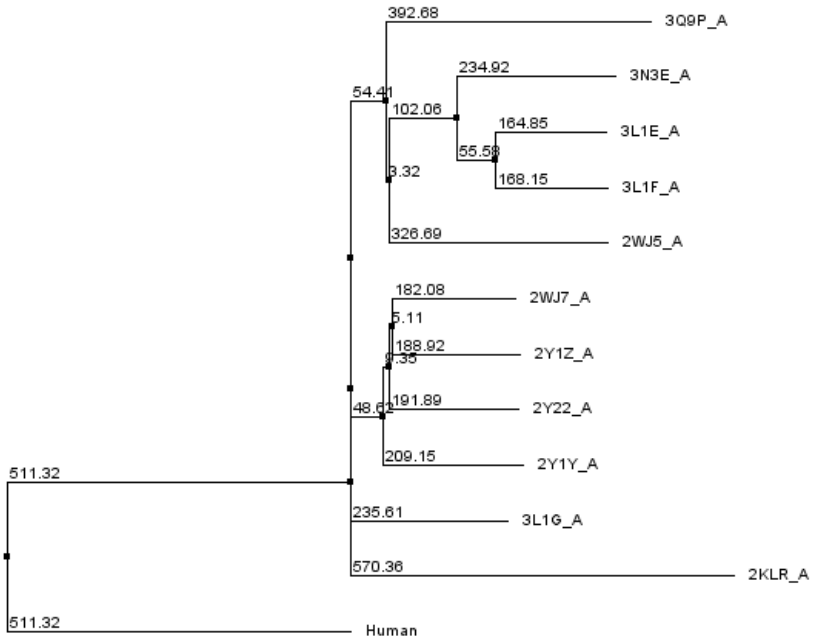

Figure 6.7: Phylogenetic tree of the small heat shock proteins generated by Jalview

Given by the fact that all of the established hits cover not the whole sequence of the heat shock protein 27, they form a great group which is only related to the query sequence. Some of them contain only the alpha-crystallin domain like in the case of 2Y1Z, 2Y22 or 2KLR. The last one has also the highest evolutionary distance in comparison to the Hsp27. 3L1G is a alphaB-crystallin structure, one of the true heat shock proteins in the human, and separated from the other sequences. All of the other hits are classified into two groups. The upper group contains the alpha-crystallin domains of several chordata species like rat (2WJ5), bovine (3L1E, 3L1F) and zebrafish (3N3E). Completed is these group by the sequence of the human alpha-crystallin domain. As second group are the sequences of the alphaB-crystallin. It approved the relation of the different heat shock proteins and the closer relation within the heat shock protein sequences in the single groups.

6.4 Functional Placement of the Heat Shock Protein 27

6.4.1 Functional Placement by Gene Ontology

Depending on the increased amount of data which is stored in several databases, there are different terminologies used to describe the data. On this basis it is difficult to ex-

change the developed knowledge. Hence, some formal and explicit specifications of the terms are established and the different relationships between them have been defined. These specifications are called ontologies[54]. The widely used collaborative project is called Gene Ontology (GO) and describes genes and gene-associated information for all organisms[55]. A GO is defined by a number and a name which is then called a GO term. Each term is categorized to one of the subontologies "molecular function", "cellular component" and "biological process"[55]. The GO terms with the three subontologies belonging to Hsp27 are shown in table A.1. Given by the fact of the great size the table has, it would displace the formatting at this point and is listed in the appendix A.

The GO terms of the subontology "molecular function" suggest that the protein Hsp27 binds to the protein kinase c. This kinase is an enzyme which phosphorylates other proteins by transfer a phosphate group, mostly ATP, onto a protein substrate. It is inactivated by the heat shock protein 27 which also binds ubiquitin. Ubiquitin is a protein which marks other proteins for the proteolytic degradation by bind them covalently.
All cellular components are established by the correspondent GO term. In this case are all components listed which are related to the heat shock protein 27 by containing this protein or cause modifications of this component when Hsp27 is expressed on a higher level.
Behind the subontology "biological process" all processes are collected in which Hsp27 is involved. It is involved into the RNA and the mRNA metabolism like every protein caused on the fact that a protein is a DNA product and is developed by different steps like transcription and translation. Also the gene expression is a process which stands in such a relation to the Heat shock protein 27. The angiogenesis is approved by the KEGG database. This biological process is established as the blood vessel formation from the pre-existing blood vessel cells to create new blood vessels. In combination with this biological process stands the second process blood vessel endothelial cell migration. This movement from an endothelial cell into the extracellular matrix takes place when new blood vessel cells are developed. As most important fact about the heat shock protein 27 is the protection of other cells against the apoptosis which is also listed in the GO terms as anti-apoptosis and cell death. The last term defines when a cell is declared as death. Therefore, three criteria are described and one of this criteria must be conformed to declare a cell as death. Hsp27 is also involved in the cellular component movement but given by the fact that the cell itself is declared as cell component, this GO term appears for every protein. Comparable to this process is the endothelial cell chemotaxis which describes the direct movement of an endothelial cell which is controlled by a specific chemical concentration gradient. The heat shock protein 27 regulates several processes in a negatively way and therefore inhibits or stops these processes. To these processes count the apoptosis, the phosphorylation, the protein kinase activity and the serine or threonine kinase activity. Especially the last two processes are listed before under the molecular function. But Hsp27 regulates some processes also in a positive way. The interleukin-1 beta production and the tumor necrosis factor biosynthesic pro-

cess are regulate positive by Hsp27. Only the regulation of the I-kappaB kinase or the necrosis factor (NF)-kappaB cascade and the translational initiation are regulated by Hsp27 but the way is not defined. The heat shock protein 27 is expressed in response to stress, unfolded proteins and viruses which cause state or activity changes of the cell.

6.5 Association of mutations and known disease patterns

There are 17 polymorphisms of Hsp27 known but only 8 of them cause two disease pattern. The first disease pattern is the distal hereditary motor neuropathy type IIB (HMN2B), the second disease is the axonal Charcot-Marie-Tooth disease type 2F (CMT2F). Both disease patterns will be described later but first the characterization of the mutations will be illustrated.
An overview of the polymorphism is given in table 6.7 corresponding with the disease patterns, the conservation and the secondary structure elements (SSE).

ID	Polymorphism	Disease Pattern	Conservation	SSE
1	Val6Phe	no disease pattern	0.0	not annotated
2	Gly13Ala	no disease pattern	0.0	not annotated
3	Ser15Ile	no disease pattern	1.0	not annotated
4	Glu41Asp	no disease pattern	0.0	not annotated
5	Leu58Met	no disease pattern	0.0	not annotated
6	Pro60Ser	no disease pattern	0.0	not annotated
7	Gln80Lys	no disease pattern	1.0	not annotated
8	Leu99Met	HMN2B, autosomal recessive	2.0	B-strand
9	Leu109Arg	no disease pattern	2.0	B-strand
10	Thr121Ser	no disease pattern	2.0	B-strand
11	Arg127Trp	HMN2B	9.0	beta-hairpin
12	Ser135Phe	CMT2F, HMN2B	8.0	beta-hairpin
13	Arg136Trp	CMT2F	4.0	B-strand
14	Arg140Gly	HMN2B	5.0	B-strand
15	Thr151Ile	HMN2B	3.0	H1-helix
16	Pro182Leu	HMN2B	7.0	not annotated
17	Pro182Ser	HMN2B	7.0	not annotated

Table 6.7: Summary of the known polymorphism of Hsp27

For understanding the significance of the single mutations, it is necessary to have a detailed view on the associated amino acids. Figure 6.8 shows the amino acids and their different side chains for better appreciation of the following statements.

The first polymorphism is an exchange of the amino acid valine at position 6 against

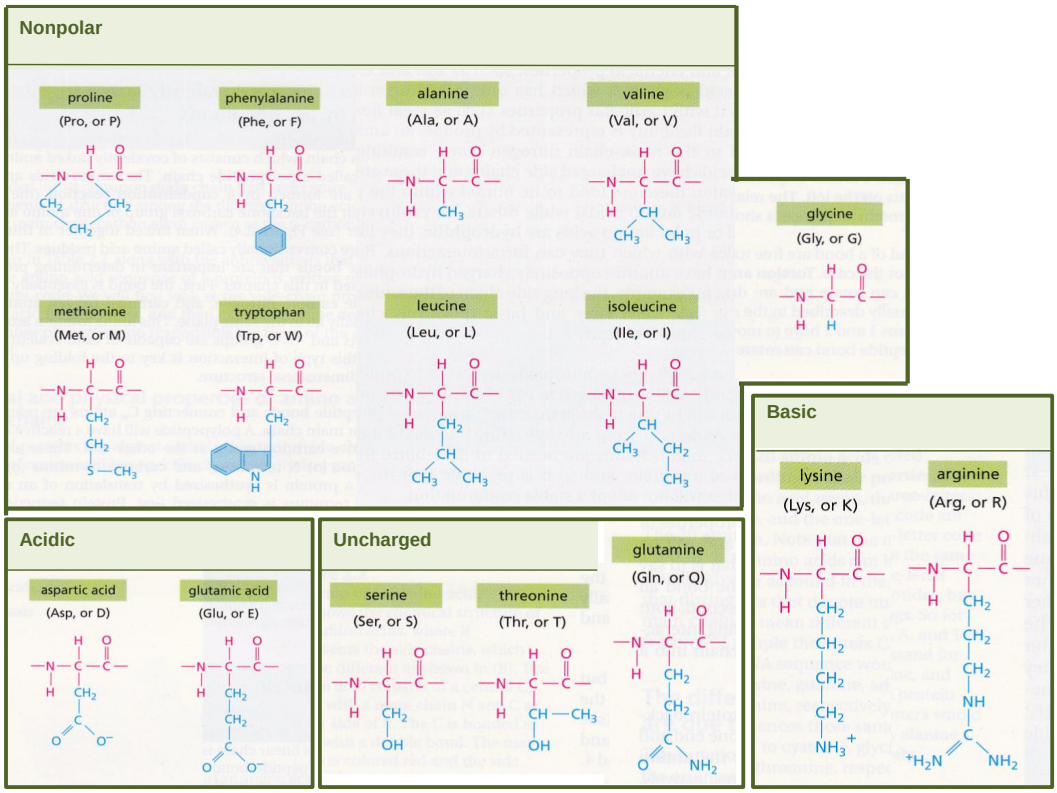

Figure 6.8: The associated amino acids with their different side chains arranged by their chemical properties[53]

the amino acid phenylalanine. Both side chains are nonpolar but valine only contains a methylethyl group whereas the side chain of phenylalanine owns a benzyl ring. The aromatic compound causes a bigger size of the side chain, given by the fact that a benzyl ring is very stable in its conformation based on its electron distribution. These size changes can obtain structural modifications in the native conformation of the protein structure. However, the polymorphism is located in a region without any conservation which is probably a reason for the missing relation to a disease pattern.

As second polymorphism an exchange at position 13 is listed where a glycine is substituted by an alanine. Like in the polymorphism before, both amino acids own nonpolar side chains. Glycine is the smallest amino acid because it contains only one hydrogen as side chain. Therefore, alanine is also a tiny amino acid and owns only a methyl group in its side chain. This exchange has probably no large consequences for the protein structure because the size of the involved amino acids differs not crucial. Above all a conservation is not given in this region.

The third polymorphism is an exchange of the amino acid serine against isoleucine which takes place at position 15. The conservation in this region is with 1.0 a little bit higher then by the polymorphism before but it is not high enough to declare it as a conserved area. Serine is an amino acid with an uncharged side chain which contains a hydroxyl group. These chemical group is highly reactive and obtains interactions to other reactive groups and the belonging molecules. In comparison, isoleucine contains

a side chain with a methylpropane group which declares this amino acid as a nonpolar one. This exchange causes probably a structural conformation change of the protein depending on the size differences of the two amino acids. Given by the fact that the hydroxyl group of serine is deleted by the exchange, the number of interactions will decrease and this will also cause structural changes.
As fourth polymorphism the exchange of glutamic acid at position 41 against aspartic acid is listed. Both have acidic side chains and the single difference is a hydrocarbon group. This results in a size change because glutamic acid is the bigger amino acid and can obtain a structure modification of the native protein structure. An reactive difference between these two amino acids should not exist because both own a carboxyl group with a free oxygen atom. The region is also not conserved.
The fifth polymorphism results in an exchange of the amino acid leucine at position 58 against methionine. Their properties are similar, they have both a nonpolar side chain, but methionine has a sulfate in its side chain. Hence, it would be possible to create disulfide bonds and therefore change the structural properties of the amino acid and the belonging neighbors. Given by the fact that the region is not conserved, no relations to disease pattern exists. The secondary structure element is not annotated which contains these polymorphism. It is possible that the exchange of the two amino acids cause functional modifications but this can not be proven until the structure of the protein is enlightened.
At position 60 the amino acid proline is exchanged against serine. Proline is the only nonpolar amino acid which integrates the aromatic ring into its amino group. This cause a great size of the amino acid which obtains in an exchange to a conformation change. Serine has an uncharged side chain with owns a hydroxyl group. Therefore, the aromatic ring is exchanged by a highly reactive functional group which cause more interactions and a smaller size then the aromatic compound. The region of the polymorphism is not conserved and no disease pattern is related.
The seventh polymorphism is the exchange of the amino acid glutamine at position 80 against the amino acid lysine. Thereby is an uncharged side chain replaced by a basic side chain. The free proton of the amino group in the side chain of the lysine cause more interactions with the environment of these reactive group. This property and the greater size of the replaced side chain results in a conformation change of the protein structure. However, the polymorphism appears in a region of no conservation.
The eighth mutation results in an exchange of the amino acid leucine at position 99 against methionine which is described above at position 58. Their similar properties are results of the same nonpolar side chain, only methionine owns a sulfate in its side chain and can associates stronger bonds with other sulfate groups which obtain more interactions. This similarity in chemical characteristics can be the reason that the disease pattern only occurs in humans who having this mutation in both chromosomes. Therefore, it is passed on to an autosomal recessive that only homozygous suspenders have the disease.
As ninth polymorphism the exchange of the amino acid leucine at position 109 against arginine is listed. Leucine has a nonpolar side chain with a propyl group which will not

associate strong bonds to other molecules because it is saturated itself and has no free atoms which are able to develop some bonds. In comparison, arginine has a basic side chain with a guanidine group which makes this amino acid solvent in water. Guanidine is a derivative of carbamide or urea[25]. The difference in the charge of the side chains can cause conformation changes when the amino acid arginine is incorporated instead of leucine. The region of this exchange is therefore in a low conservation which hold the consequences on a common level. Hence, this mutation is not associated with a disease pattern.
The tenth polymorphism takes place at position 121 and exchanges the amino acid threonine against serine. Threonine has an uncharged side chain with a ethoxy radical group which is a very small side chain. The hydroxy group is really unstable and therefore very reactive to get a saturated electron distribution. The amino acid serine has also an uncharged side chain but obtains a methoxy radical group. This little difference in the size of the side chain should not cause large consequences because it is also located in a region of low conservation. Therefore it is not associated to a disease pattern or other abnormalities.
As eleventh mutation the exchange of arginine at position 127 against tryptophan is listed. The side chain of arginine is basic so it exists a positive charge caused by the amino group which is bind additionally. By the exchange against tryptophan this positive charge is removed and a nonpolar side chain appeared for it. The nonpolar side chain contains an aromatic compound which makes it possible to associate more binding sites and more interactions now. This fact and the fitting of a bigger amino acid could disturb the native conformation of the protein so the disease pattern occurs. Additionally, the mutation is in a region of high conservation what can cause terrible functional changes.
The twelfth mutation is an exchange of serine at position 135 against phenylalanine. Serine carries an uncharged polar side chain and is a very tiny amino acid. In comparison, phenylalanine has a nonpolar side chain and contains an aromatic compound. Therefore, with the new amino acid phenylalanine it is possible to get more interactions and the native conformation is changed by the bigger amino acid. Also in this case is the mutation located in a region of high conservation where changes have large consequences for the normal protein function.
As thirteenth mutation the amino acid at position 136, arginine, is exchanged by tryptophan. It is the same exchange like in the second mutation but the region is not such highly conserved.
The fourteenth mutation in sequence is an exchange at position 140 and replaces the amino acid arginine by glycine. Arginine has a basic side chain and leads to a positive charge. By exchanging the amino acid by glycine the conformation could change because glycine is a very small amino acid. It is located in a region with middle conservation so the consequences could not be so large like in a higher conservation.
As fifteenth mutation the amino acid threonine is exchanged by isoleucine at position 151 in the sequence. Threonine has an uncharged polar side chain containing a hydroxyl. In comparison, isoleucine carries a nonpolar side chain which contains more methyl groups then threonine so more interactions are possible which would change the

native conformation of the protein structure.
The last two mutations are located both at position 182 where a proline is exchanged. Proline is the only amino acid where an aromatic ring is involved in the amino and the carboxyl group. However, it carries a nonpolar side chain and is a small amino acid. It is exchanged by leucine and serine. Leucine is also an amino acid with a nonpolar side chain and has some more methyl groups which could get new interactions. Serine is a very tiny amino acid and carries an uncharged polar side chain. In both exchanges the native conformation would change and this would have large consequences in a region with high conservation. The following figure 6.9 shows the described mutations in the protein structure model of Hsp27.

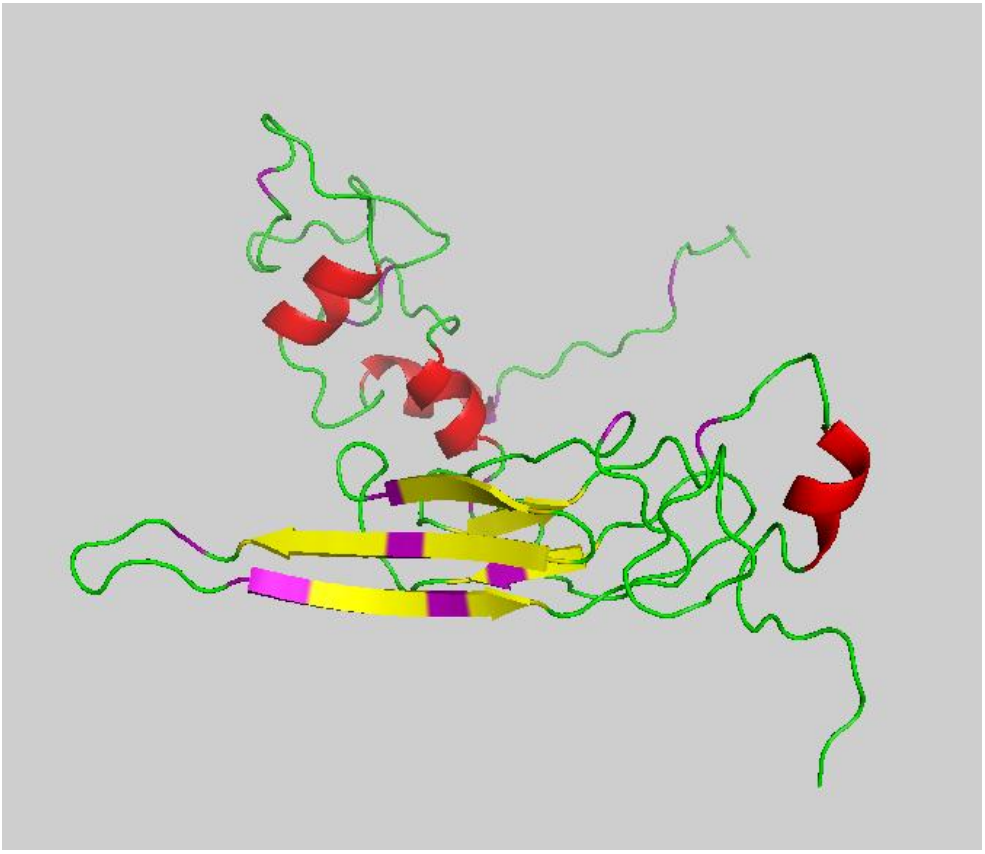

Figure 6.9: The established mutations colored in purple are displayed in the generated structure model

The known mutations cause two disease patterns, distal hereditary motor neuropathy type 2B and the Charcot- Marie-Tooth disease as type 2F. Both diseases are caused by mutations in the location 7q11.23 where also the gene for the Hsp27 is localized. CMT2F is an allelic disorder and has a similar phenotype like HMN2B[36]. Additionally, of the HMN2B, three variations exist - HMN2A, HMN2B and HMN2C. All are caused by mutations of small heat shock proteins[36]. HMN2B was first reported 2008 where 5 families showed a remarkably similar slowly progressive disease[36]. The first symp-

toms are muscle weakness and atrophy started in the distal lower limb muscles followed by muscle weakness and wasting to the upper limbs 5 to 10 years after the disease pattern appeared[36]. As affection point of time could be detected from the ages 21 to 54[36]. However, sensory abnormalities could not be detected in all cases, but tendom reflexes are depressive[36]. Decisive for the disease are 4 mutations in the HSPB1 gene and most of them are heterozygous[36]. Only the mutation at the position 99 is homozygous. Although the phenotype of the two disease patterns are similar no HMN2B mutations are found in 90 families with CMT2F[36].
It was reported in 2001 that a 6-generation family with autosomal dominant CMT2F occurred in a province of Russia[35]. The disease occurred at the age of 15 to 25 and shows progressive weakness just as atrophy of the lower limb muscles like by HMN2B[35]. In the case of CMT2F the atrophy and weakness causes foot drop and steppage gait. When the muscles in the upper limb also starts to become weak it causes clawing on the hands[35]. At early stage the deep tendom reflexes are depressive or absent[35]. All patients have mild or moderate sensory impairments in the feet and hands[35]. CMT2F is a slowly progressive disease, resulting in disability after 15 to 20 years without loss of fitness or restricted life spam[35].

Summarized the two disease patterns have similar phenotypes which are the result of the mutations taking place at the same location. The main part of the mutation is located within the alpha-crystallin domain of the Hsp27. This domain is a protein binding site and a highly conserved region. Hence, it is possible that this area is important for the protein structure and function. Therefore, little changes could have large consequences for the maintenance of the native protein structure. Hsp27 is known to be involved in signal transduction processes which can obtain the disease pattern. The weakness of the muscles is related with neuronal transmittance of the signals. This cause a loss of the movement and therefore misapplication of the muscles which achieve the degradation of the muscle cells. It is not the aim to declare this thesis as incontrovertible but it is a possible reaction chain for the disease pattern.

7 Structure Modeling of Hsp27

In this chapter the modeling process of the structure belonging to the heat shock protein 27 will be described. The aim is to generate a complete structure model and to analyze in the 9th chapter the binding possibilities of the ligand RP101.

7.1 Description of the Used Software Tools

7.1.1 SwissModel

If SwissModel is used for the modeling each user has three possibilities of application. The first variant is the "Automated Mode" which generates a model automatically based on different templates stored in the common databases like Protein Data Base[23]. As guideline the sequence similarity should reach 50% as minimum[23]. To start the automated modeling it is required to set the protein sequence or the UniProt ID as input[23]. The templates for the model are chosen by a internal BLAST run and are primarily generated with a high resolution[23].
In the "Alignment Mode" a multiple sequence alignment (MSA) is integrated and should be used as a starting point for comparable modeling[23]. It is important that for one member of the MSA the tertiary structure must be known[23]. With the alignment possibility different alignments could be tested and the quality of the alignments is proved and assesed[23].
The "Project Mode" combines the two previous methods and guarantees a better control[23]. As input a so called project file is necessary which contains every important information and orders for the modeling[23].

7.1.2 Protein Structure Prediction Server

The protein structure prediction server combines the PSI-BLAST, IMPALA and T-coffee as effective consensus strategy[10]. It establishes a template selection and fitting template-target alignments[10]. By using MODELLER the tertiary structure of the target is generated[10]. It is possible to modulate the template selection and chose between PSI-BLAST or IMPALA[10]. Also a PDB template and for the modeling process the RAMP method could be used[10].

7.1.3 CPH models

The template identification is based on profile-profile alignments and orientated on secondary structure as well as adjournment prediction[34]. As input it is only the sequence required[34]. For the modeling process only the best fitting template with the highest identity is used[34].

7.1.4 Pcons

The software Pcons evaluates the quality of different models and sorts them on the basis of their assessment[51]. For the template models several structures are used as input[51]. They are generated with the condition that the models of each template was established by different methods and approaches[51]. Pcons is also able to search for repeated three dimensional structure cluster which are assessed by a score how common this motives are in relation to the whole template set[51]. It acts on the assumption that a pattern which appears more often has a higher probability to be correct[51]. There are two scores generated, one for the global quality and one for the local quality where each residue is a basis[51].

7.1.5 M4T

M4T is an approach which uses a multiple mapping method with multiple templates and results in a selection and a combination of the most suitable templates[14]. It applies an iterative clustering method which takes account to the sequence similarity to the target and among the templates[14]. Also the completeness of the structural domains is important and its experimental quality are involved into the mapping method[14]. For the template search is the database PDB used so all templates are declared with the PDB IDs.

7.1.6 I-TASSER

The server could predict the protein structure and therefore the possible function[2]. Is is based on multiple-threading alignments by LOMETS and repeated TASSER assembly as simulation[2]. The structure is adjusted with a protein function database[2]. I-TASSER has the aim to generate a most accurate structural and functional prediction with the most modern methods[2].

7.1.7 ModWeb

ModWeb is a web server for automated comparative modeling based on PSI-BLAST, IMPALA and MODELLER which declare similarities to the protein structure prediction server who uses also the first two approaches[32]. As input are one or more FASTA sequences accepted and the models are calculated from the best template structures found in the PDB[32]. It is also possible to deliver a protein sequence as input data and identify sequence homologous by the SwissProt database[32].

7.2 Analysis of Different Templates for Structure Modeling

7.2.1 Template Search for the Modeling Process

The "Automated Mode" of the SwissModel establishes two possible templates for a structure modeling process. These two templates are 3Q9P_A and 2YGD_T which also appears in the sequence analysis of the protein Hsp27. Their quality is assessed by the SwissModel and is summarized in the table 7.1.

Criterion	**3Q9P_A**	**2YGD_T**
Sequence identity	91.358%	40%
E-value	3.70125e-32	1.1e-36
QMEAN Z-score	1.323	-5.048
Quaternary structure	monomer	single chain
Amino acids	90-170	17-196

Table 7.1: Summary of the found templates by SwissModel

On the basis of the lower Z-score for the template 2YGD_T SwissModel advises against the usage of this template for a modeling process. Also the sequence identity is with 40% under the necessary threshold for an automated modeling project generated by SwissModel.
The disadvantage of 3Q9P_A as template is the gap about 8 amino acids appearing in the coil region 125-132 of the protein structure. Therefore, a covered region of 85 amino acids is not enough to generate a protein structure model because the N-terminal and C-terminal sequence is missing in the template. In comparison, 2YGD_T has a bigger range of protein structure at its disposal but only a sequence identity of 40% with several coil regions. Hence, it could be combined with the 3Q9P_A template when the modeling process generates no sufficient structures.

The template search by the protein structure prediction server established 4 hits identified by the PSI-BLAST run. As the best hit with a score of 81 and a sequence identity of

17% the protein structure 2BOL_A is generated which is the metazoan small heat shock protein Tsp36. This structure model is deposit in the PDB with sulfate ions as ligands which probably change the native conformation of the quaternary structure. The next two established templates are 1GME_A and 1GME_D belonging to the eukaryotic small heat shock protein Hsp16.9. They differ in the sequence identity. Hence, the chain A has a higher identity of 18% while chain D has only 16%. Also the last template 1SHS_A which is the PDB ID for a small heat shock protein of *Methanococcus jannaschii* exhibits a low sequence identity of 17%. However, all 4 established templates have a low sequence identity in a range of 16% to 18% which could make the modeling process difficult.

The software CPH models detected 3Q9P as the best fitting template based on the high sequence identity of 91.4%. As second hit the template 2YGD is established like by the SwissModel before. Afterward some new templates are listed which are also established by Pcons and M4T. The score is very low so Pcons advises against all generated templates because the results are not trustworthy enough. To simplify the evaluation of the templates they are summarized in the following table 7.2. With the exception of the template 3GLA, the other templates are established by the PSI-BLAST in the previous chapter.

PDB ID	**Description**	**Sequence Identity**	**Occurrence**
3L1G	alphaB-crystallin	56%	CPH, Pcons
3N3E	zebrafish alpha-crystallin	54%	CPH, M4T
2WJ7	alphaB-crystallin	53%	CPH
2WJ5	rat alpha-crystallin domain	56%	CPH, Pcons
2Y1Z	alphaB-crystallin, alpha-crystallin domain	50%	CPH
2KLR	alpha-crystallin domain	45%	CPH
2Y22	alphaB-crystallin domain aa 67-157	52%	CPH
3L1E	Bovine alphaA-crystallin Zinc Bound	54%	Pcons, M4T
3GLA	crystal structure of the hspA from *Xanthomonas axonopodis*	42%	Pcons

Table 7.2: Summary of the established templates by Pcons, CPH model and M4T

ModWeb has generated 21 models which could be used as templates for a modeling process. Some of these models appeared manifold but the region which should be used as template differs. The following table 7.3 shows all established models with the belonging template regions. Each model is assessed by the E-value and three scores. For the main part of the generated models is the sequence identity too low for a usage by a structure modeling process. Only the models of 3N3E_A, 2WJ7_A and 3Q9P_A should be used as templates. They reach a sequence identity of 50% and better, but they cover only the region of the alpha-crystallin domain. All templates which covers the left over regions offer such low sequence identities that a modeling process is not commendable.

PDB ID	Template Region	Sequence Identity	Sufficient Assessment
2BOL_A	6-170, 8-205, 15-205	15%, 18%, 18%	E-value
1GME_A	12-168, 26-184, 46-185	22%, 22%, 13%	E-value, GA341
1WH0_A	74-199	9%	E-value, Z-score
1RLR_A	82-182, 82-185	13%, 11%	E-value, Z-score
3LJY_A	82-200	11%	E-value, Z-score
1GME_B	83-182, 87-184, 87-184	20%, 19%, 23%	E-value, GA341, Z-score
3N3E_A	84-188	50%	recommended for modeling
2JKI_S	84-177	6%	E-value, Z-score
1SHS_A	85-182, 87-181	both 17%	E-value, GA341, Z-score
2WJ7_A	86-170	56%	recommended for modeling
3Q9P_A	90-171	99%	recommended for modeling

Table 7.3: The established models of ModWeb

Therefore, a combination of the good templates with some templates which contain a lower identity but fit to another important region of the query sequence.

7.2.2 Structure Modeling by I-TASSER

I-TASSER generates a predicted secondary structure for the input sequence of the Hsp27 and is assessed by a confidence score. The following table 7.4 shows the amino acid regions with the belonging predicted structure element and the established score as mean value because the confidence score is calculated for each amino acid.

The main part of the predicted structure is created by coil structures. Within the region of the alpha-crystallin domain the expected sheet structures appear. Six sheet structures are also established for the secondary structure of the HSPB1 fragment. For the gap in the structure of 3Q9P is a combination of coil and sheet established by I-TASSER. Some helix structures appear in the prediction but they are mostly short.
I-TASSER also generates a solvent accessibility which appears in a range of 0 (buried) to 9 (highly exposed)[2]. Most of the structure is not solvent accessible so these regions will be located inside the molecule and have no interactions to other components. After the amino acid 160 a higher exposed region begins so maybe it would be located on the surface of the molecule to interact with other molecules like the ligand RP101.
I-TASSER lists the used templates and shows the alignments of these sequences. As expected it is the highest conservation and best accordance within the alpha-crystallin domain. The templates are the dame like established by the template search before. For the structure modeling 2YGD_A, 2BOL_A and 1GME_A are used. The five predicted models for the protein structure of the Hsp27 are sorted by a established C-score. This score is generated out of the significance of the alignments called Z-score and the convergence of the I-TASSER. It lays in a range of -5 to 2 and stands for a better quality as higher this score is. On the basis that the first model has a score of -0.34 it would be

Region	Predicted SSE	Score
1-26	coil	5.69
27-32	helix	3
33-38	coil	7.5
39-43	helix	6.8
44-87	coil	4.25
88-89	sheet	2
90-94	coil	5
95-100	sheet	8.16
101-105	coil	4.8
106-107	helix	3.5
108-109	coil	4.5
110-114	sheet	8
115-116	coil	9
117-127	sheet	7.5
128-132	coil	5.8
133-143	sheet	8.45
144-149	coil	8.16
150-152	helix	7.33
153	coil	8
154-157	sheet	7.5
158-161	coil	8.25
162-168	sheet	7.28
169-179	coil	7.09
180-184	sheet	5
185-205	coil	4.8

Table 7.4: The predicted secondary structure for the query sequence

used as the best predicted model. The overlay of the five generated models shows that they are very similar and have only little differences in the structures. Only the coil structures change in their orientation and arrangement but not in a way to differ the function in a greater aspect.

The large number of coil elements open the possibility that the generated model is not equally to the native conformation of the protein structure. An analysis by energy profiles of the sequence authenticates these thesis. Therefore some additional sequence analysis should be realized to establish some new information for the modeling process. First the generated coil structures are used as input for the PDB and structures with the sequence motifs are searched. For the most coil structures no other sequence motifs could be found.

Some sequence alignments with the detect structure templates have been made. The two templates 3Q9P_A and 1GME_A overlap in a region of 36 amino acids which reach the previous part of the alpha-crystallin domain. The program Jalview owns the property

to calculate a secondary structure prediction which shows similar structure elements in the overlapping area of the templates. Some predicted beta-strands have almost the same length. Also the addition of other generated templates can not give more information which could be integrated into the structure prediction.

A possibility for another structure prediction is the ROBETTA server. This server has different features, interesting for the project are the structure prediction and the fragment library. The last one divide the sequence into triplets and search in a database for these triplets the equally sequence motifs. Their belonging secondary structure is stored and evaluated. By a relation to the PSIPRED program this triplets are collected to a whole secondary structure prediction which is validated by a confidence score like used before by I-TASSER. The structure prediction feature need only the sequence as input and calculate therefore the predicted structure information. Unfortunately, this calculation failed by generating over four weeks without any status change, so it is stopped after this time.

7.3 Molecular Analysis of the Generated Structure Model

With the help of the visualizing program PyMol the structure of the heat shock protein 27 is analyzed. The following figure 7.1 shows the established secondary structure with the transparency surface. The right small picture is colored in the electrostatic potential of the sequence. For a better orientation the model with the normal surface is laid underneath. In the left small picture the structure is visualized with some highlighted point which stands for the potential polar contacts. They are sites for possible binding areas.

The right picture shows the electrostatic potential energy of the three dimensional molecule and visualizes the charge distributions on the surface. These distribution strikes the observer because the largest part of the protein structure is blue and therefore a positive charge. Only little parts appear in a red color and adverts to the negative charge. With the exception of the C-terminal region of the structure are the coil elements negative charged. Whereby the secondary structure elements like strand and helix are positive charged. This facts are necessary to establish possible binding sites for the docking of the ligand RP101. The detailed information about the RP101 are collected in the following chapter.
The polar contacts describe possible binding sites for hydrogen bonds. The largest number of polar contacts are detected within the secondary structure elements strands and helices. Reasonable for this effect is the conformation of the structure elements. Therefore, the strands are created by hydrogen bonds between the single primary structures. In the helix elements these hydrogen bonds are required for the rotation the primary structure achieves and which creates the special conformation.

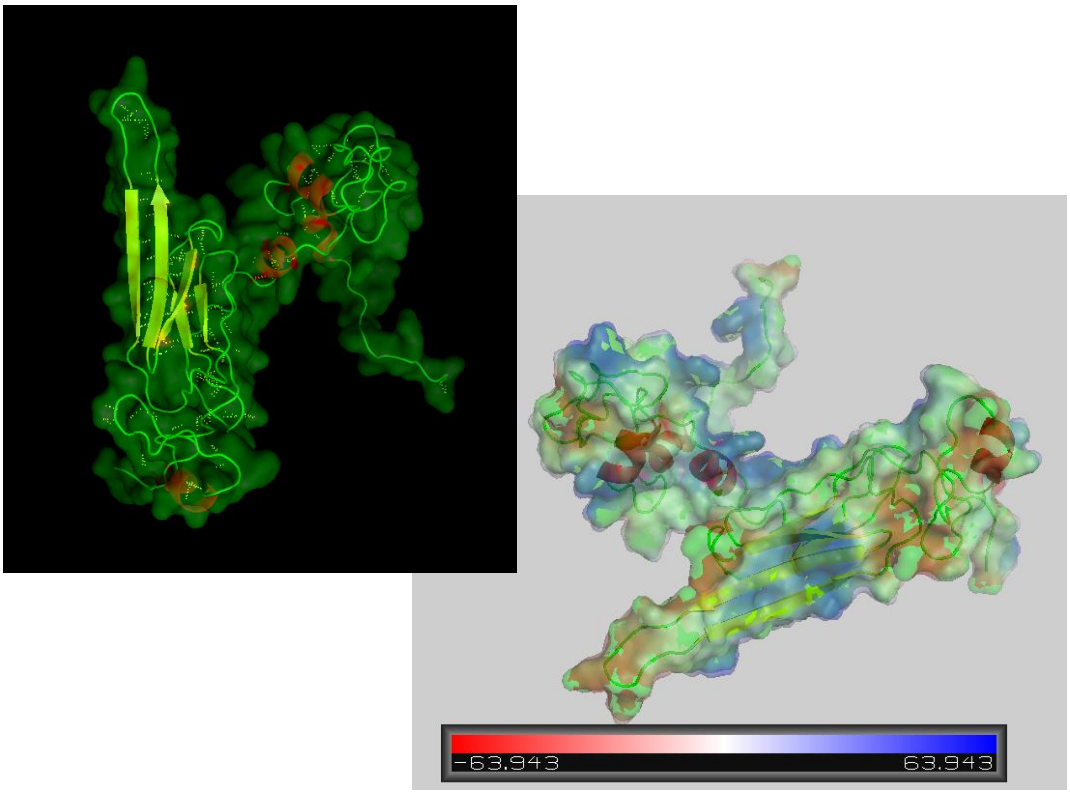

Figure 7.1: Polar contacts and the electrostatic potential

The solvent accessibility, generated by the program Swiss-PDBViewer, is visualized in the figure 7.2.

The color scheme ranges from blue (buried) to red (highly exposed). The whole structure of the heat shock protein 27 is colored in yellow and green which suggest that the main part of the protein is not highly exposed but also not totally buried. Only the region within the structure is light blue so there are residues collected which are not solvent. As single red spot appears the C-terminal end of the structure which is therefore highly exposed. This fact is caused on the free carboxyl group which is located at the end of the last residue. Therefore, the solvent accessibility is used for the secondary structure prediction and collects information for the movement of a molecule depending on its environment.

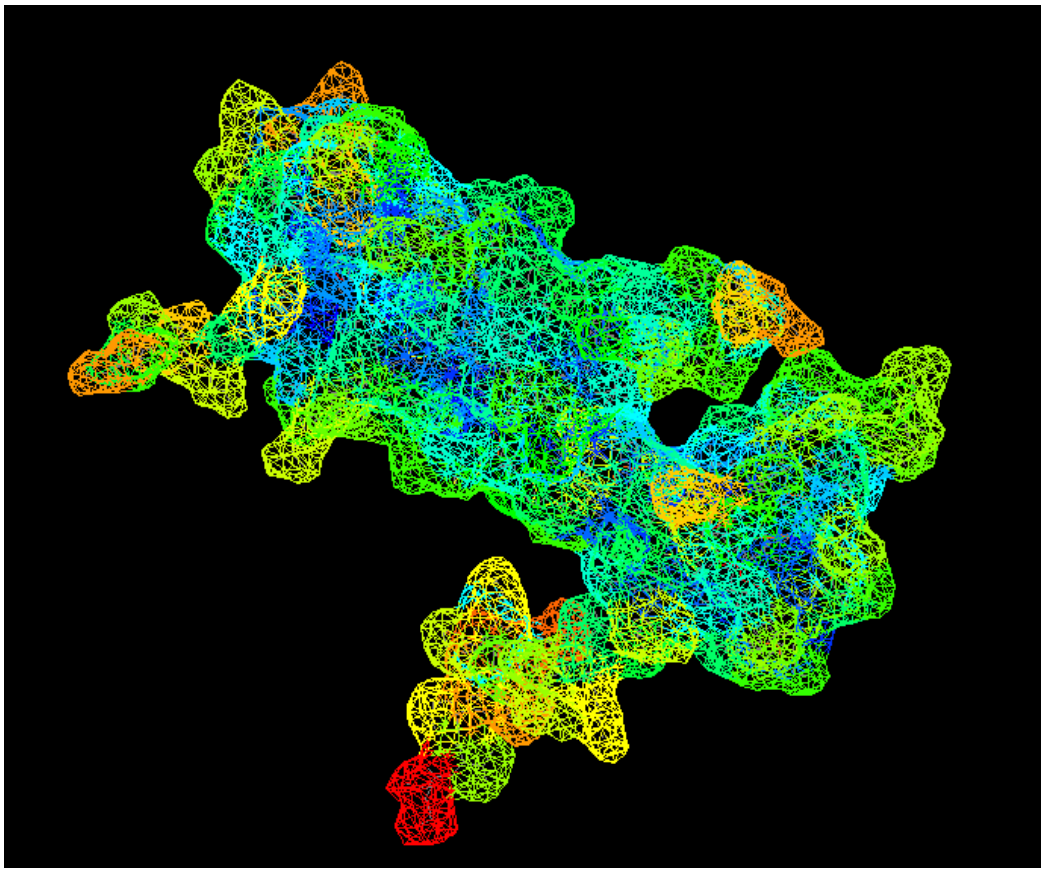

Figure 7.2: The solvent accessibility of the predicted structure of the Hsp27

8 Analysis of the Ligand RP101

This chapter contains the analysis of the ligand RP101 which is a abbreviation for the biochemical compound (E)-5-(2-Bromvinyl)-2'-deoxyuridine. In the pharmacy it is also known under the name brivudine abbreviated by BVDU.

8.1 General Information about RP101

RP101 is a thymidine anti-metabolite and inhibit DNA sequences. Its chemical properties are summarized in the following table 8.1.

Property	**Value**
CAS number	69304-47-8
PubChem ID	446727
molecular formula	$C_{11}H_{13}BrN_2O_5$
molecular weight	333.135g/mol
density	1.857g/cm^3
melting point	165°C
boiling point	382.9°C

Table 8.1: Chemical Properties of the Compound Brivudin

Brivudine is an antiviral drug, effective against herpes simplex virus type 1 and varicella zoster virus. The last one causes herpes zoster, also known as shingles. Comparable pharmaceutica are Aciclovir, Valaciclovir and Famciclovir. RP101 is distinguished by a higher antiviral potency and a larger half-live. It was developed 1976 in England[50] and in East Germany where it was used for the treatment of herpes simplex and herpes zoster[48, 52].

8.2 Mechanism of Action in Virus Infected Cells

The potency of BVDU has been demonstrated by various clinical studies[3, 4]. This potency depends on the specific mechanism of the action which is described in the following explanations.
BVDU is phosphorylated by the virus-encoded thymidine kinases (TK) for two times. Depending on the virus which infect the cells, the herpes simplex virus 1 TK or the varicella zoster virus TK cause the phoshorylation. Hence, brivudine converts to its 5'-monophosphate and 5'-diphosphate[11]. The phosphorylation step to the 5'-triphosphate is done by a nucleoside 5'-diphosphate kinase. This triphosphate can interact with the

viral DNA polymerase in two ways. First it is able to be used as a competitive inhibitor to the natural substrate dTTP. Second it could be an alternative substrate which allows the incorporation of the triphosphate into the growing DNA chain[1]. By this incorporation the stability and the function of the DNA is affected during the replication and the transcription process. Hence, the viral DNA is not replicated and the resulting disease is stopped or probably even not appear. The figure 8.1 shows this process of action in a schematically way.

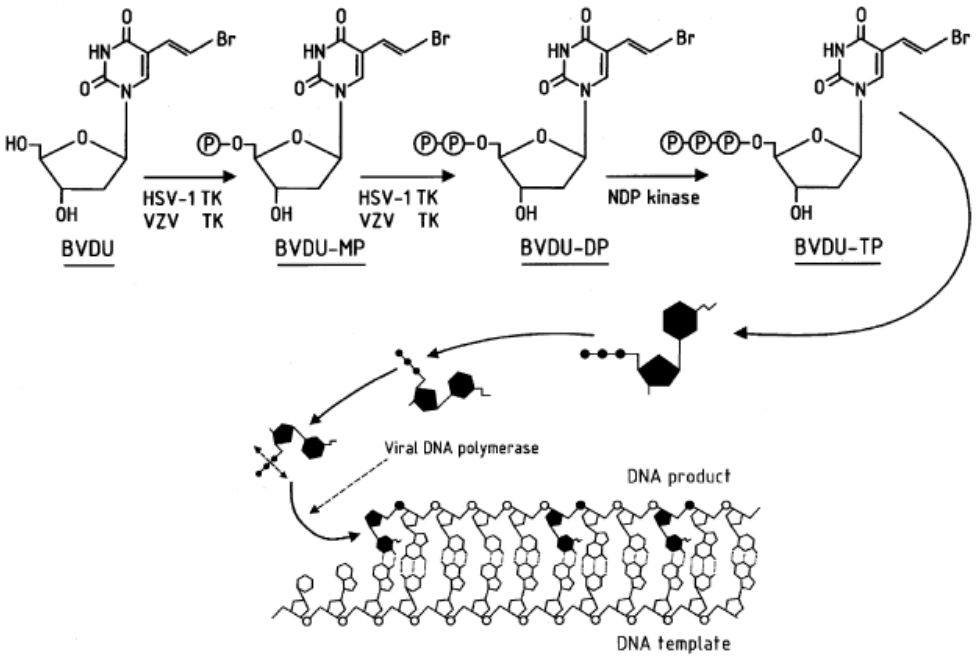

Figure 8.1: RP101 Mechanism of action in virus infected cells

This presented mechanism of action will not be used in the case of the inhibition of the heat shock protein 27. Reasonably for this fact is the loss of the two specific thymidine kinases which cause the first two phosphorylation steps. These two necessary reaction steps can not be developed by other kinases. Hence, the triphosphate as reactive compound is not created and therefore no incorporation into the DNA chain is possible.
The inhibition of the heat shock protein 27 must be realized by another way of action, possibly by covalent binding. In this case the native conformation can be changed so the active center will be shifted and is not able to sustain its biological function anymore. It is also possible that the chemical compound RP101 does not bind onto the active center and change the protein structure conformation by binding to another region of the protein surface. To check this effect it is necessary to analyze the molecular and chemical properties of the surface of brivudine.

8.3 Analysis of the Molecular and Chemical Properties of RP101

With the molecular visualizing program PyMol the three dimensional structure of the RP101 is analyzed. Given by the fact that it is a thymidine anti-metabolit, RP101 has a uridine nucleoside as basis. This is a chemical component consisting of uracil bind to a ribofuranose. Uracil can develop two hydrogen bonds to the corresponding nucleoside base, adenine. The hydrogens, bound to the nitrogen, are able to cater to other atoms like oxygen, nitrogen and fluorine. These hydrogen bonds are very strong and create a stable binding to the molecule which contains the atom for the hydrogen bond. All atom bonds are saturated, so every free oxygen or nitrogen has bound a hydrogen. Therefore, no additional bonds could be developed. The surface of the molecule has a neutral charge allover, visualized in the figure 8.2. Given by the fact that the chemical compound has a neutral charge, an analysis of electrostatic potential can offer no more information.

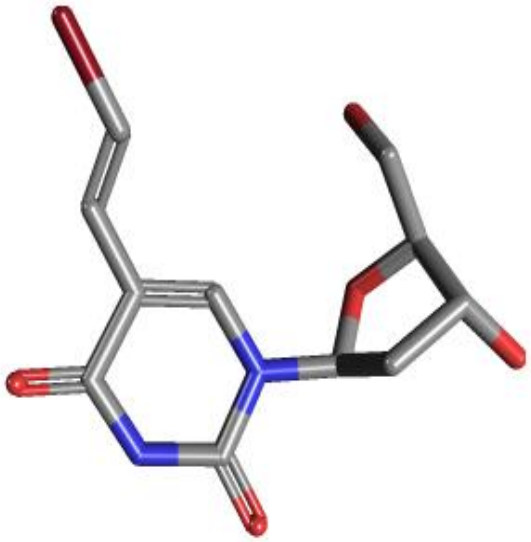

Figure 8.2: Brivudine in the three dimensional structure

9 Docking of the Structure Model with the Ligand

This chapter describes the docking process which should combine the generated structure model with the ligand RP101 to identify the position within the sequence where the ligand is bound.

9.1 Establishment of Potential Binding Sites for Hsp27 and RP101

The docking process should be generated by the program HADDOCK which generates the protein-protein complexes by combining the energetic approach with shape complementary[49]. In the most cases for such calculations exists no experimental data or no clear information of the docking process could be developed.
But first, it is necessary to generate possible binding sites for the involved molecules. These sites are then used as input for the most docking programs to narrow the calculation time down. By using the previous molecular visualizing programs no binding sites could be detected. Hence, it was necessary to establish the binding sites by other programs.

Related to the docking program HADDOCK the program naccess generates such binding sites. It must be installed on a computer and will run on a windows system by using a virtual machine if no unix system is usable. By the establishment of the needed input files and the necessary course, the condition appeared that the molecule should be declared as protein or nucleic acid. Given by the fact that RP101 is a chemical compound no binding sites can be generated by using the program naccess. Hence, another usable program has to be found for the discovery of the potential binding sites. The web server Q-SiteFinder, available under http://www.modelling.leeds.ac.uk/qsitefinder/[27], generates a specific number of binding sites which are chosen by the user.
The established ten binding sites for the structure model of Hsp27 are written in a Pymol file. This file can be viewed by the molecular visualizing program Pymol and declares the generated binding sites in the protein structure. The belonging figure 9.1 shows the location of these binding sites. Each single binding site is a potential binding site also if the coloring scheme is similar they are not defined as one.

As similar program, Pocket-Finder (http://www.modelling.leeds.ac.uk/pocketfinder/) generates the same binding sites for the heat shock protein 27. An enhancement of the two previous programs is the server LIGSITEcsc which collects several programs to establish potential binding sites. This server authenticates the predicted binding sites from

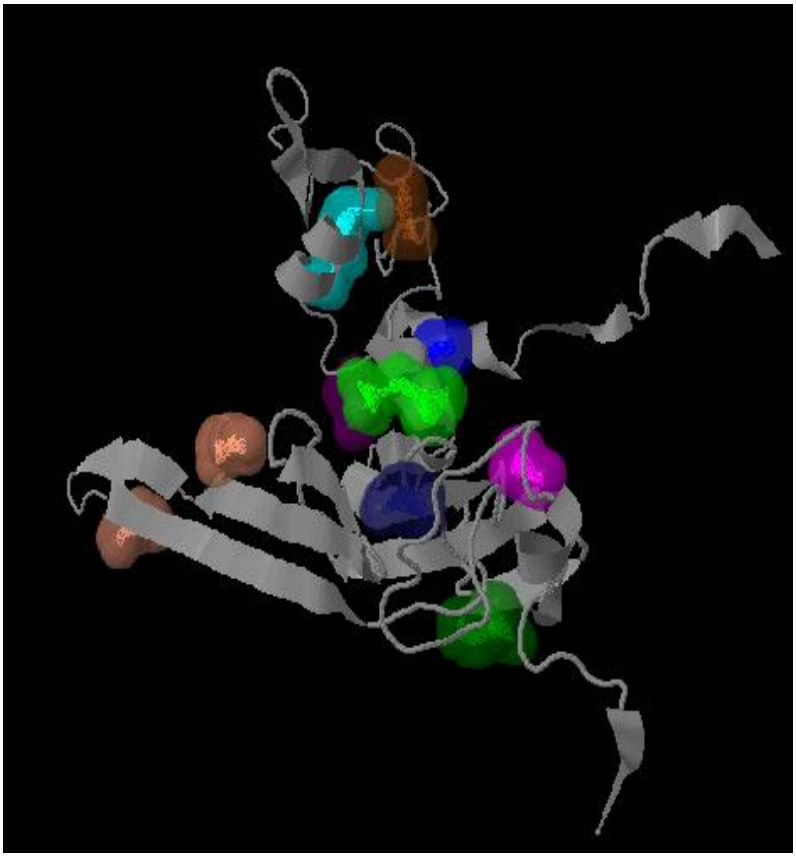

Figure 9.1: Binding sites generated by the program Q-SiteFinder

the figure 9.1 caused by the fact that Q-SiteFinder and Pocket-Finder are integrated into LIGSITEcsc.
For the chemical compound brivudine no binding sites could be established by LIGSITEcsc or the other programs. This can be caused on the fact that all possible free atoms for a potential bond are saturated by hydrogens. Therefore, the docking process should be started without binding sites of the RP101. It is also possible that HADDOCK can not be used for the docking process if no possible aberration can be found to start the process without the necessary binding sites of RP101. It can possible that a molecular dynamic simulation can identify more molecular properties of the ligand and the structure[19], but for this approach is not enough time.

9.2 Docking Process

As the docking process should be started by using the web service of HADDOCK a problem appeared. It only combines proteins and nucleic acids, but RP101 is a chemical compound and therefore a docking generated by Haddock is not possible anymore. Hence, other possibilities for the docking should be searched.

The server SwissDock predict the molecular interactions between a target and small molecules. It is based on the docking software EADockDSS which generates the predicted docking process by the following steps. Most of the binding modes are established by local or blind docking[16]. The local docking is generated in a box, therefore the blind docking establish the binding modes in the neighborhood of all target cavities[16]. At the same time are the CHARMM energies estimated on a grid and then the most advantegous binding modes are evaluated by FACTS and clustered[16]. The most favorable clusters are established as outputs[16]. These cluster are equally to the predicted binding sites and therefore the docking areas are calculated. For using SwissDock as docking program different variants are possible. The user has the selection between the option to load a file which contains the molecule up on the server or chose some allocated files from the server itself. It is also possible to deliver the PDB ID of a protein and the ID of the ZINC database for chemical compounds. For the small molecules which are used as ligands in the docking process some conditions in file formats are determined. The necessary files are stored in the related ZINC database in different variations. The docking process itself only takes a few minutes but this probably depends on the involved molecules. As output file a chimera file can be downloaded and modulated with the chimera program user-defined. The figure 9.2 shows the result of the docking algorithm with the ligand RP101 located within the protein structure of Hsp27.

The ligand brivudine is located in the middle of the protein structure and appears near to the beta-sandwich of the alpha-crystallin domain. Given by the fact that this domain is a polypeptide binding site the resulting binding of the ligand is a certainty which has to be expected. This area is also detected by Q-SiteFinder as possible binding site. The following figure 9.3 shows a closeup view of the binding area between the target and the ligand.

Figure 9.2: The docking result of the target sequence of Hsp27 with the ligand RP101

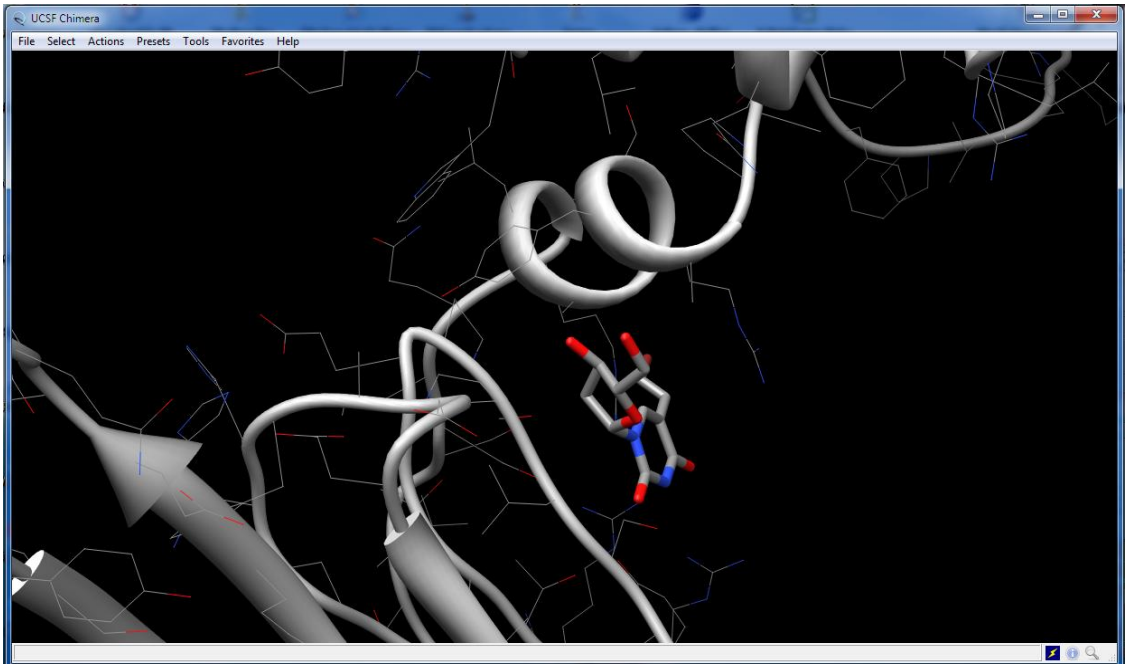

Figure 9.3: The binding area between target and ligand

10 Conclusion

Given by the fact that the expression of the titin-Hsp27-construct has failed, the structure of the protein could not be predicted by evaluating the atomic force microscopy measurements. Possibly modifications which can solve the expression problem are given in the discussion of the first part of this thesis. The prediction of the structure by using bioinformatic methods achieves its aim and a structure model is established. Unfortunately, this generated structure can not be approved by experimental data. Nevertheless, the model gives an impression how the ligand RP101 can probably bind to the heat shock protein 27.

The time period of five month determine the possibilities for the structure prediction as well as the necessity to use free available software. With a greater time period for the prediction it is possible to generate a molecular dynamic simulation to optimize the structure model and check different molecular or chemical properties. The development of the necessary files and the adaption of the simulation are very time-consuming, especially when different programs should be tested.

The development of the protein structure of Hsp27 seems to be a little bit difficult depending on the fact that the protein can not be crystallized. Therefore, the structure annotation by X-ray analysis can not be developed. This is probably a reason that until the current point of time no complete structure of Hsp27 is deposit in any database although it has such an important function and usage for the cancer therapy.

All established structure information within these master thesis can be refined by some additionally approaches like the responded molecular dynamic simulation or the usage of some other prediction software.

Appendix A: Tables and Figures

GO number	GO name	GO subontologie
0005515	protein binding	molecular function
0005080	protein kinase c binding	molecular function
0008426	protein kinase c inhibitor activity	molecular function
0019901	protein kinase binding	molecular function
0043130	ubiquitin binding	molecular function
0030018	Z disk	cellular component
0009986	cell surface	cellular component
0043292	contractile fiber	cellular component
0005737	cytoplasm	cellular component
0005856	cytoskeleton	cellular component
0005829	cytosol	cellular component
0005626	insoluble fraction	cellular component
0005622	intracellular	cellular component
0005634	nucleus	cellular component
0005886	plasma membrane	cellular component
0000502	proteasome complex	cellular component
0005625	soluble fraction	cellular component
0005819	spindle	cellular component
0016070	RNA metabolism process	biological process
0001525	angiogenesis	biological process
0006916	anti-apoptosis	biological process
0043534	blood vessel endothelial cell migration	biological process
0008219	cell death	biological process
0006928	cellular component movement	biological process
0035924	cellular response to vascular endothelial growth factor	biological process
0035767	endothelial cell chemotaxis	biological process
0010467	gene expression	biological process
0016071	mRNA metabolism process	biological process
0043066	negative regulation of apoptotic process	biological process
0042326	negative regulation of phosphorylation	biological process
0006469	negative regulation of protein kinase activity	biological process
0071901	negative regulation of protein Ser/Thr kinase activity	biological process
0032731	positive regulation of interleukin-1 beta production	biological process
0042535	positive regulation of TNF biosynthetic process	biological process
0043122	regulation of I-kappaB kinase/NF-kappaB cascade	biological process
0006446	regulation of translational initiation	biological process
0006950	response to stress	biological process
0006986	response to unfolded proteins	biological process
0009615	response to virus	biological process

Table A.1: GO terms belonging to the heat shock protein 27

Appendix B: Chemical Composition of used biochemical agents

B.1 SDS separating gel

Solution Component	Pipetting Volume
Deionized water	2.1ml
30% acrylamide mix	0.5ml
1.0M Tris (pH 8.6)	0.38ml
10% SDS	0.03ml
10% ammonium persulfate	0.03ml
TEMED	0.003ml

Table B.1: Chemical composition of the separating gel with a volume of 3ml

B.2 SDS resolving gel

Solution Component	Pipetting Volume
Deionized water	5.3ml
30% acrylamide mix	2.0ml
1.0M Tris (pH 8.6)	2.5ml
10% SDS	0.1ml
10% ammonium persulfate	0.1ml
TEMED	0.008ml

Table B.2: Chemical composition of the resolving gel with a volume of 10ml

B.3 Lysis Buffer

Concentration	Chemical Compound	Molar Mass
50mM	Sodium dihydrogen phosphate	137.99g/mol
300mM	Sodium chloride 58.44g/mol	
10mM	Imidazole	68.08g/mol

Table B.3: Chemical composition of the lysis buffer

This solution is adjusted with sodium hydroxide to a pH value of 8.0 and filled up to 1l with deionized water. For the lysis process 10ml are used and stowed with 10mg of lysozyme from the hen egg white (Fluka Biochemika) and one tablet of the complete mini EDTA-free (Roche Diagnostics).

B.4 Washing Buffer

Concentration	Chemical Compound	Molar Mass
50mM	Sodium dihydrogen phosphate	137.99g/mol
300mM	Sodium chloride	58.44g/mol
20mM	Imidazole	68.08g/mol

Table B.4: Chemical composition of the washing buffer

This solution is adjusted with sodium hydroxide to a pH value of 8.0 and filled up to 1l with deionized water.

B.5 Elution Buffer

Concentration	Chemical Compound	Molar Mass
50mM	Sodium dihydrogen phosphate	137.99g/mol
300mM	Sodium chloride	58.44g/mol
250mM	Imidazole	68.08g/mol

Table B.5: Chemical composition of the elution buffer

This solution is adjusted with sodium hydroxide to a pH value of 8.0 and filled up to 1l with deionized water.

B.6 Dialysis Buffer

The dialysis buffer is composed of 1l phosphate buffered saline, prepared by the medium kitchen of the Biotec, stowed with 0.1mM EDTA.

B.7 SOC medium

The medium was prepared by the medium kitchen of the Biotec.

B.8 2xYT medium

The medium was prepared by the medium kitchen of the Biotec.

This solution is adjusted with sodium hydroxide to a pH value of 7.0 and filled up to 1l with deionized water.

Concentration	Chemical Compound
5g/l	Yeast extract
20g/l	Tryptone
0.6g/l	Sodium chloride
0.2g/l	Potassium chloride
10mM	Magnesium chloride
10mM	Magnesium sulfate
20mM	Glucose

Table B.6: Chemical composition of the SOC medium

Concentration	Chemical Compound
900ml	Deionized water
16g/l	Yeast extract
10g/l	Tryptone
5g/l	Sodium chloride

Table B.7: Chemical composition of the 2xYT medium

Bibliography

[1] Allaudeen HS et al. (1981). On the mechanism of selective inhibition of herpesvirus replication by (E)-5-(2-bromovinyl)-20-deoxyuridine. Proc Natl Acad Sci USA, 78: 2698702

[2] Ambrish R, Alper K, Yang Z (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 5: 725-738

[3] Andrei G et al. (1992). Comparative activity of various compounds against clinical strains of herpes simplex virus. Eur J Clin Microbiol Infect Dis, 11 14351

[4] Andrei G et al. (1995). Comparative activity of selected antiviral compounds against clinical isolates of varicella-zoster virus. Eur J Clin Microbiol Infect Dis, 14 31828

[5] Altschul SF (1990). Basic Local Alignment Search Tool. J. Mol. Biology, 215: 403-410

[6] Arrigo AP (2007). Hsp27 (HspB1) and aB-crystallin (HspB5) as therapeutic targets. FEBS Letters 581: 36653674

[7] BLAST picture. Available on 01-08-2012 from: http://4.bp.blogspot.com/_8S467kTrM/9//2eAd5M6stw0/1600/BLAST_algorithm.jpg

[8] Bruey JM et al. (2000) Hsp27 negatively regulates cell death by interacting with cytochrome c. Nat. Cell Biol., 2: 645652

[9] CATH database, ID 2.60.40.790. Available on 07-08-2012 from: http://www.cathdb.info/cgi-bin/search

[10] Chen CC, Hwang JK, Yang JM (2006). (PS)2: protein structure prediction server. NAR, 34: W152-W157

[11] Descamps J, De Clercq E (1981). Specific phosphorylation of E-5-(2-iodovinyl)-20-deoxyuridine by herpes simplex virus-infected cells. J Biol Chem 256: 59736

[12] Dialysis picture. Available on 15-06-2012 from: http://irfanchemist.files.wordpress.com/2009/04/dialisis2.jpg

[13] Electroporation picture. Available on 21-02-2012 from: http://www.btxonline.com/product_images/uploaded_images/electroporation-cartoon.png

[14] Fernandez-Fuentes N et al. (2007). M4T: a comparative protein structure modeling server. NAR, 35: W363W368

[15] Garrido C (2006). Heat shock proteins 27 and 70. Cell Cycle, 5(22): 2592-2601

[16] Grosdidier A (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. NAR, 39: W270-7

[17] Growth curve picture. Available on 27-06-2012 from: http://gcserevision101.files.wordpress.com/2009/02/bacteria-growth.jpg

[18] Hansen A (2004). Bioinformatik. Ein Leitfaden für Naturwissenschaftler. 2. überarbeitete und erweiterte Auflage. Birkhäuser Verlag, 54-62

[19] Heinrich JC (2011). RP101 (brivudine) binds to heat shock protein HSP27 (HSPB1) and enhances survival in animals and pancreatic cancer patients. J Cancer Res Clin Oncol, 137: 13491361

[20] Hütt MT, Dehnert M (2006). Methoden der Bioinformatik. Springer-Verlag Berlin Heidelberg, 190-195

[21] Kappé G et al. (2004). Tsp36, a tapeworm small heat-shock protein with a duplicated alpha-crystallin domain, forms dimers and tetramers with good chaperone-like activity. Proteins, 57(1):109-17

[22] Kato K et al. (1994). Dissociation as a result of phosphorylation of an aggregated form of the small stress protein, hsp27. J. Biol. Chem., 269: 1127411278.

[23] Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Research. 37: D387-D392.

[24] Klemenz R (1991). Alpha B-crystallin is a small heat shock protein. Proc Natl Acad Sci U S A., 88(9): 36523656

[25] Latscha HP, Kazmaier U, Klein HA (2008). Organische Chemie. Chemie-Basiswissen II. Sechste, vollständig überarbeitete Auflage, Springer Verlag, 146-196

[26] Labudde D (2011). Molecular Modeling and Drug Discovery. Beschreibung der Struktur von Biomolekülen. 16.10.2011: slide 16

[27] Laurie AT, Jackson RM (2005). Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics, 21: 1908-1916

[28] Launay N (2006) Cell signaling pathways to alphaB-crystallin following stresses of the cytoskeleton. Exp. Cell Res. 312: 35703584.

[29] Mancini WR, De Clercq E, Prusoff WH. (1983). The relationship between incorporation of E-5-(2-bromovinyl)-20-deoxyuridine into herpes simplex virus type 1 DNA with virus infectivity and DNA integrity. J Biol Chem, 258: 7925

[30] McEntyre J, Ostell J, editors (2002-). The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21101/

[31] Merkl R, Waack S (2009). Bioinformatik Interaktiv. 2. erweiterte und neubearbeitete Auflage. Wiley-Blackwell, 219-222

[32] Narayanan E et al. (2010). Tools for comparative protein structure modeling and analysis. NAR, 31(13): 3375-3380

[33] Needleman SB, Wunsch CD (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biology, 48: 443-453

[34] Nielsen M., Lundegaard C., Lund O., Petersen TN (2010). CPHmodels-3.0 - Remote homology modeling using structure guided sequence profiles. Nucleic Acids Research, 38: W576-W581

[35] OMIM database, ID 606595. Available on 23-07-2012 from: http://omim.org/

[36] OMIM database, ID 608634. Available on 23-07-2012 from: http://omim.org/

[37] PANTHER Classification System, ID PTHR11527(:SF33). Available on 07-08-2012 from: http://www.pantherdb.org/panther/globalSearch.jsp?cust=Fam

[38] Pfam database, ID PF00011. Available on 07-08-2012 from: http://pfam.sanger.ac.uk/

[39] PHYLIP program package. Available on 14-08-2012 from: http://evolution.genetics.washington.edu/phylip/general.html

[40] Protein Information Resource, ID PIRSF036514. Available on 07-08-2012 from: http://pir.georgetown.edu/cgi-bin/

[41] ProSite database, ID PS01031. Available on 07-08-2012 from: http://prosite.expasy.org/

[42] Ritossa F (1962). A new puffing pattern induced by heat shock and DNP in Drosophila. Experimentia, 18: 571-3

[43] SDS-PAGE picture. Available on 16-07-2012 from: http://www.imb-jena.de/ rake/Bioinformatics_WEB/gifs/sds_page1.gif

[44] Smith TF, Waterman MF (1981). Comparative biosequence metrics. Journal of Molecular Evolution, 18: 38-46

[45] Superfamily 1.75, ID SSF49764. Available on 07-08-2012 from: http://supfam.org/SUPERFAMILY/index.html

[46] SPRINT Database, ID PR00299. Available on 07-08-2012 from: http://www.bioinf.manchester.ac.uk/cgi-bin/dbbrowser/sprint/

[47] Theriault JR (2004) Essential role of the NH2-terminal WD/EPF motif in the phosphorylation-activated protective function of mammalian Hsp27. J. Biol. Chem. 279: 2346323471

[48] Töpke H et al. (1984). Erste Erfahrungen mit dem Antiherpetikum (E)-5-(2-Bromvinyl)-2'-Desoxyuridin (BrVUdR) bei der Behandlung des Herpes corneae. Folia ophthal, 9: 39-48

[49] de Vries SJ et al. (2010). The HADDOCK web server for data-driven biomolecular docking. Nature Protocols, 5: 883-897

[50] Walker RT et al. (1978). The synthesis and properties of some antiviral nucleosides. NAR 1036. special publication no 4

[51] Wallner B, Elofsson A (2006). Identification of correct regions in protein models using structural, alignment and consensus information. Protein Sci., 15(4):900-913

[52] Wutzler P et al. (1988). BVDU-Therapie von Zostererkrankungen bei Patienten mit malignem Grundleiden. Z gesamte inn Med, 43: 67780

[53] Zvelebil M, Baum JO (2008). understanding Bioinformatics. Garland Science, 30

[54] Zvelebil M, Baum JO (2008). understanding Bioinformatics. Garland Science, 54-55

[55] Zvelebil M, Baum JO (2008). understanding Bioinformatics. Garland Science, 348

[56] Zvelebil M, Baum JO (2008). understanding Bioinformatics. Garland Science, 570-574

Declaration of Originality

I hereby declare that this thesis and the work reported herein was composed by and originated entirely by me.

Information derived from the published and unpublished work of others has been acknowledged in the text and references are given in the bibliography.

This thesis has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education.

Mittweida, 23 August 2012