Computational Search of the Interaction between Melanopsin & Cryptochrome proteins

by

Evrim Besray Ünal

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science
in
Computational Sciences and Engineering

Koc University

November 2006

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Evrim Besray Ünal

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Burak Erman, Ph. D. (Advisor)

_____

I. Halil Kavaklı, Ph. D. (Co-Advisor)

_____

Atilla Gürsoy, Ph. D.

_____

Alkan Kabakçıoğlu, Ph. D.

_____

Özlem Keskin, Ph. D.

Date: _____

To my Family

ABSTRACT

Circadian rhythms are oscillations in the biochemical, physiological, and behavioral functions of organisms that occur with a periodicity of approximately 24 hours. In mammals, circadian rhythm is generated by a molecular clock. The molecular clock, which is located at suprachiasmatic nuclei (SCN) part of brain, is synchronized by environmental light-dark cycle.

The cryptochromes are the mammalian circadian photoreceptors; they absorb light and transmit the signal to the molecular clock. The cryptochromes and melanopsin (and possibly other opsin family pigments) have been proposed as circadian photoreceptor pigments that exist in the inner retina. Experimental studies imply that there is most probably an interaction between melanopsin and cryptochromes for molecular clock to function normally. In order to uncover this interaction; the tertiary structures of Melanopsin and Cryptochrome; the possible interaction between those two proteins have been predicted by usage of different computational means.

The results of this study imply that mammalian Melanopsin and Cryptochrome proteins interact. The N-termini of Cryptchrome protein interacts with C-termini and cytoplasmic tails of Melanopsin protein. The in vivo interaction is supported by preliminary flourescent microscopy technique.

ÖZET

Canlıların biyolojik, fizyolojik ve davranışsal fonksiyonlarındaki salınımlara Sirkadyan ritmleri adı verilir. Sirkadyan ritmler yaklaşık olarak 24 saatlik döngülerle tekrarlanırlar. Memelilerde bu ritmler içsel bir saat tarafından düzenlenir; iç saat mekanizması beyinde bulunan suprakiazmatik çekirdek (SCN) adlı yapı tarafından kontrol edilmektedir. SCN, çevresel ışık-karanlık evreleri ile senkronize çalışır.

Retina tarafından algılanan ışık SCN'ye optik sinirler aracılığı ile haber verilmektedir. Retinada ışık algılayıcı proteinler bulunmaktadır. Bu proteinler retinadaki ışığı algılayıp içsel saate sinyal iletimi yapmak ile görevlidirler. Cryptochrome ve Melanopsin adlı proteinler (muhtemelen opsin familyasından başka proteinler ile birlikte) Sirkadyan ritmlerde görev alan, ışığı algılayan fotoreseptörlerdir.

Deneysel sonuçlar iç saatin düzenli çalışması için Cryptochrome ve Melanopsin proteinlerinin muhtemel bir etkileşim içinde olmalari gerektiğini göstermektedir. Bu çalışmada, olası etkileşimi kanıtlayabilmek için Melanopsin ve Cryptochrome proteinlerinin tersiyer yapıları; ve iki protein arasındaki etkileşim hesaplamalı yöntemler kullanılarak bulunmuştur.

Çalışmanın sonuçları memeli Melanopsin ve Cryptochrome proteinlerinin etkileşim içerisinde olduğunu göstermektedir. Cryptochrome proteinin N-ucu Melanopsin proteinin C-ucu ve sitoplazmik kısımları ile etkileşmektedir. In vivo etkileşim, flüoresan mikroskopi tekniği ile desteklenmektedir.

ACKNOWLEDGEMENTS

The easiest and hardest chapter for me to write; naming all the people who helped me will be easy; but I will not be able to thank them sufficiently with words.

I would like to thank, first and foremost, my advisors Prof. Burak Erman and Assistant Prof. Halil Kavaklı for their guidance and support throughout my graduate study and during the completion of this thesis. Without their encouragement, knowledge and enthusiastism this thesis would have never been completed.

I am extremely grateful to Assistant Prof. Alkan Kabakçıoğlu, Assistant Prof. Atilla Gürsoy and Assistant Prof. Özlem Keskin for their participation in my thesis committee and for the critical reading of my thesis.

I would like to thank Assistant Prof. Alper Kiraz and Şennur Turgut for collaboration.

I would like to thank Dr. Şule Özdaş for her extreme helps in my laboratory studies and for encouraging me all the time. I would like to thank my research group friends Aytuğ Tuncel, Emre Özdemir and Natali Özber for supporting me and for their cheerful friendships.

I would like to thank all my friends at Koç University who made the two years an enjoyable period of time; Ayşegül, Canan, Cem, Emre, Hazal, Nesrin, Nurcan, Miray, Osman, Suat and especially Deniz, Özge, Selen, Zeynep. I am grateful to Özkan who accompanied me in all good and bad times.

Last, I would like to thank my grandparents İhsan & Kadri Gamsızlar; Emine & Mahmut Ünal; my parents Işın & Osman; my sister Eren Naz for their love, affection and support throughout my life, without their encouragement I could not be what I am today. To them I dedicate this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

A variety of physiological processes of animals, plants, fungi and cyanobacteria are determined by endogenously generated rhythms; called *Circadian Rhythms*. Chronobiologist Franz Halberg is the creator of the word “Circadian”; *circa* means “around” and *dies* means “day” in Latin, so the word means *around/about a day*.

Circadian clock of human is the oscillations in the biochemical, physiological and behavioural patterns in approximately 24 hours period. The rhythm is intrinsic, and regulated by light-dark cycles. The light signal is detected by eyes; the signal is transmitted to suprachiasmatic nuclei (SCN) part of brain; this signal is interpreted and regulation of the organisms’ behaviour and physiology is achieved by secretion of hormones.

The perception of light and signal transduction cascade is not known for sure. There are only some experimental evidences and hypothesis about the signal transduction pathway. Those hypotheses all indicate that opsin family proteins and/or cryptochromes should play role in signal transduction pathway.

The cryptochromes, which were first discovered in Arabidopsis, are the blue-light photoreceptors. They absorb light and transmit the signal to blue-light dependent of signal transduction.

Opsins are the transmembrane proteins that act as receptors for extracellular ligands. Those proteins are classified as classical and non-classical opsins. Classical opsins, such as rhodopsin, are known to play role in vision and colour perception. Melanopsin is a member

of non-classical opsin class. This protein is found in the inner layers of mammalian retinal cells; but it shows resemblance to invertebrate opsin proteins. Unlike classical opsins of mammals, melanopsin does not function in visualizations tasks; it was shown to act as photoreceptor.

In mammals, the cryptochromes and melanopsin have been proposed as circadian photoreceptor pigments that exist in the inner retina to transmit signal to the SCN to tell the time of day. Both humans and mice have two cryptochrome proteins; Cry1 and Cry2. Cry2 is mostly expressed in retinal cells.

Based on current evidence it is possible that mammalian Cry2 plays role in the light dependent signal-transduction via interaction with Melanopsin. We have taken both computational and experimental approaches in order to show possible interaction between them during the circadian photoreception.

To prove the hypothesized interaction of melanopsin and cryptochrome an experimental design is the best way; but experimental approaches for interaction/interactions between Cry2 and melanopsin have many shortcomings. Melanopsin is a transmembrane protein and Cry2 has FAD and MTHF as cofactors; consequently these two proteins are hard to purify at their functional states. As a result an alternative way; *computational approach* was taken in order to see whether any interaction between the two proteins occurs. Computational studies lead us to an optimum result, afterwards experiment was established.

In this study, computational biology approach was used in order to be able to model the tertiary (3D) structure of Cry protein; find the interaction between melanopsin and Cry2; to reveal the signal peptides and post-translational modification sites of those proteins.

Chapter 2 provides necessary background on circadian rhythms, melanopsin, cryptochrome and basics of computational biology tools.

Chapter 3 describes the bioinformatics tools used, in detail; the mechanisms underlying the 3D modeling and protein-protein interactions are explained. The comparison of the available tools and the most reliable tools are discussed.

Methods are given in Chapter 4. Data under various conditions are described and their results are given in Chapter 5, with the discussion of the results.

The thesis is concluded with a short summary of the performed study and future research work.

Chapter 2

OVERVIEW

2.1 Circadian Clock Overview

Circadian rhythms can be defined as the oscillations in the biochemical, physiological, and behavioral functions of organisms that occur with a periodicity of approximately 24 hours [1].

The circadian rhythms have 3 major properties;

1. Rhythms are generated intrinsically, but they are also dependent on the external stimuli, such as light-dark cycles, temperature changes. The endogenous property of the rhythms was first observed by de Marian in 1729 [51], with his study on heliotrope plant. de Mairan had shown that the daily rhythm of the plant (the movements of the leaves) was still persistent in 24-hour dark conditions.
2. Photoentrainment; rhythms can be reset/synchronized by external light-dark cycles.
3. Temperature compensation; rhythms are able to maintain their normal behaviour despite of the external temperature alterations.

The intrinsic rhythm, photoentrainment and temprature compensation properties of the circadian system are determined and maintained by different cell types, tissues and organs.

According to those different properties; the circadian system is made up of three components; the input component, the clockwork component, and the output component (Figure 2.1).

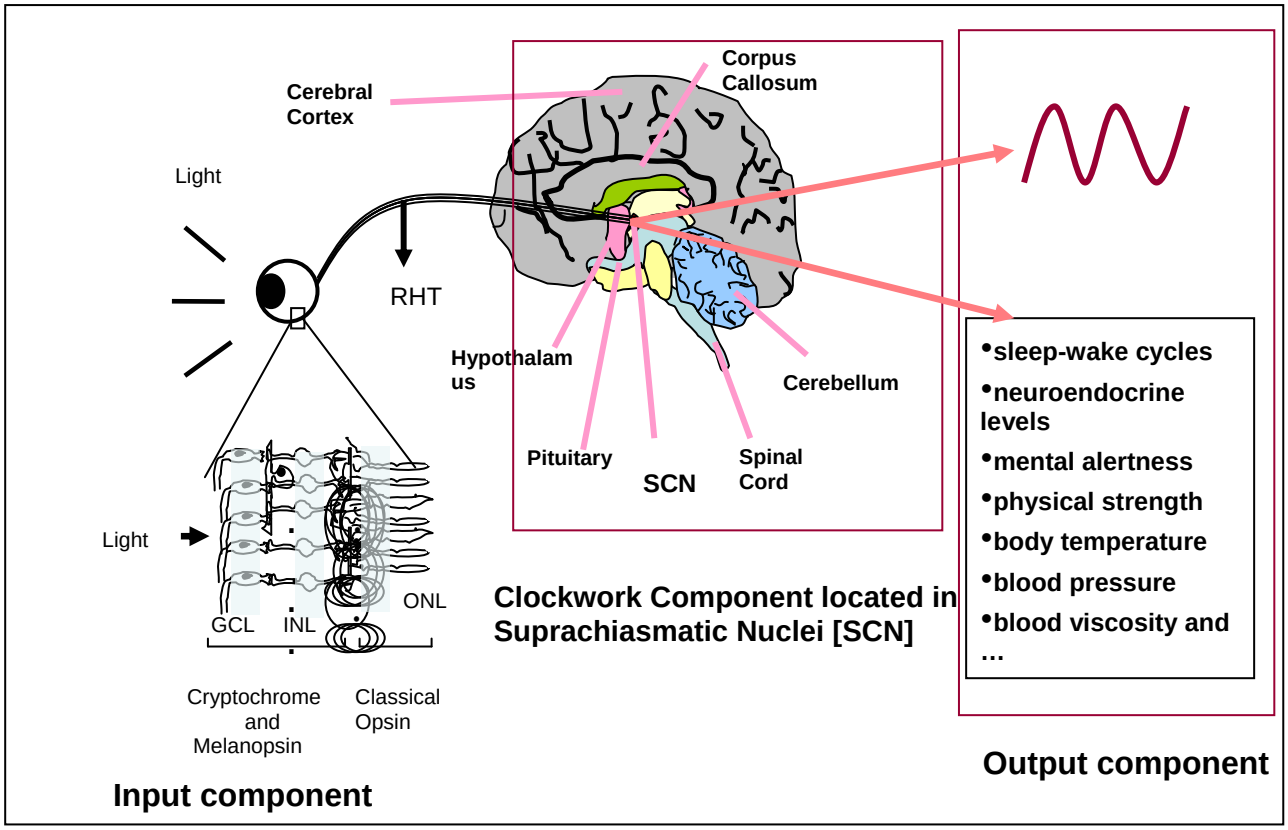


Figure 2.1: Components of the Circadian rhythm. Cryptochromes are expressed in INL (inner nuclear layer), GCL (ganglion cell layer); melanopsin are expressed in GCL [1].

2.1.1 Input Component

The input component, which we are interested in, is the perception of light. The mammalians perceive light is via their eyes. The pathway from eye to SCN (Figure 2.1) occurs both from GCL and ONL using opsin based photoreceptors. However, a wealth of

evidence suggest that cryptochrome also participates in light transmissions from eye to the SCN. According to evidence the candidates of photoreceptors are classified;

1. Non-classical opsins
2. Cryptochromes and classical opsins
3. Cryptochromes, classical opsins and non-classical opsins together.

Cryptochrome:

Cryptochromes (Cry) are blue-light receptor proteins, which display a high homology to the DNA photolyases. DNA photolyases are DNA repair enzymes, which are activated by blue-light. Apart from sequence similarity, Cry was discovered to have the chromophores that DNA photolyase has; flavin adenine dinucleotide (FAD) and 5, 10-methenyl tetrahydrofolate (MTHF). The chromophores of Cry have the ability to play role in the electromagnetic transmission of signal from eyes to SCN. Although both Crys and photolyases are similar in structural level, Crys do not posses DNA repair activity [1, 11].

Cry genes were first discovered in *Arabidopsis thaliana* via a mutation in blue-light signaling. A gene corresponding to *HY4* locus of *Arabidopsis thaliana* was isolated by Ahmad et al [2]. The inhibition hypocotyl elongation response of this plant is dependent on blue-light; and the *hy4* mutant is selectively insensitive to blue-light. The *HY4* gene was found to be encoding a protein, which is very similar to DNA photolyase. Consequently, the protein encoded by *HY4* has a structure similar to blue-light photoreceptor, it was the first Cryptochrome discovered [2].

Comparision of primary amino acid sequences of various crypthochromes and photolyases revealed that Cry posses extended C-terminal of length 50-250 amino acids [3]. Also these comparisons yield four phlogenetically different [4] families;

1. Animal cryptochromes/6-4 photolyase
2. Plant cryptochromes

3. Cry-DASH proteins
4. CPD photolyases.

The first cryptochrome in human was found when *E. coli* DNA photolyase was aligned against human EST databank [12]. It has been discovered that the human Cry (hCry) proteins show 30% similarity to E. coli DNA photolyase. The second cryptochrome of human was found in Prof. Aziz Sancar's lab [1]. The human Cry proteins (hCry) was found to have FAD and MTHF as cofactors; but they lack photolyase activity. The hCry were named as; Cry1 and Cry2 [13]. Both hCry proteins have an extended C-terminal; consisting of 100-200 amino acids. These C-tails are thought to have a significant role in functions of Cry proteins [3].

The expression of cryptochromes in many diverse organisms indicates that they have a common, vital role; they have been demonstrated experimentally to regulate the circadian clock [5]. The similarity of cryptochromes to DNA photolyases indicates that they can also be activated by blue-light. The mammalian cryptochromes are expressed in all tissue types; but highly expression is observed in tissues containing circadian photoreceptors; the gene for Cry1 is highly expressed in mouse SCN, and gene for Cry2 is highly expressed in retinal cells [6].

The studies with knock-out mice also support the idea that cryptochromes regulate clock; the Cry1 knock-out mice show a shorter biological clock-cycle, whereas the Cry2 knock-out mice have longer clock-cycle [8, 9, and 10]. Mice with both cryptochromes knocked-out have arrhythmic biological clock under constant darkness [8, 9]. In mammalians, Cry1 and Cry2 expressions oscillate with respect to the daily light-dark cycle. All of these experimental results all indicate that cryptochromes have an essential role in circadian pathway.

The most prominent circadian photoreceptor of Drosophila is found to be cryptochrome [7]. The cryptochromes are also circadian photoreceptors of *Arabidopsis thaliana* plant, and it is probable that they may be universal circadian photoreceptors [11].

Opsins:

Opsin proteins consist of 7 alpha-helices sitting in the cell-membrane; accordingly they are also named as 7-transmembrane (7TM) receptors. 7TM receptors have 3 cytoplasmic and 3 extracellular loops connecting the membrane-bound helices (Figure 2.2). Opsins play role in light-dependent signal pathways, such as vision, color perception, etc. They transmit an extracellular light signal to an intracellular electrochemical response.

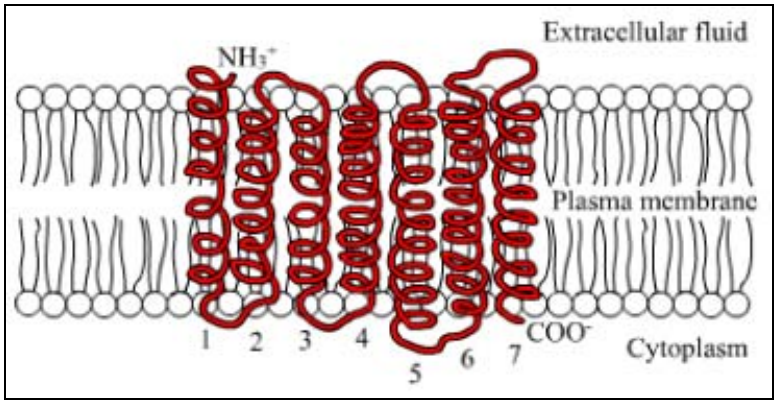


Figure 2.2: A 7-transmembrane receptor protein illustration. The alpha-helices that sit in the membrane are numbered from1 to 7.

Opsins contain a vitamin-A base chromophore named *Retinaldehyde*, also known as retinal. A retinal binding pocket exists in the transmembrane region of the opsins. Generally the chromophore is covalently bound to a Lysine residue of the 7^{th} transmembrane alpha-helix of the opsin with a Schiff-base linkage. There are 3 modes of the chromophore; 2 *cis* modes and 1 *trans* mode. When extracellular light signal is

absorbed by either 2 cis forms of chromophore; a conformational alteration occurs in the protein. This conformational change leads to accumulation of signal into the cell; and signal transduction starts. After the absorption of light, the chromophore takes the trans form and becomes inactive. The trans form is substituted by a new cis-formed chromophore for further detection of light stimuli.

Opsin family proteins are highly conserved; the Lysine residue at 296th location (Lys296) and Glu113 are conserved for the retinal linkage to the protein; Glu134-Arg135-Tyr136 residues are conserved for the signal transfer.

The opsin family proteins can be classified as; **classical** and **non-classical opsins**. Classical opsins are located in the *Outer Nuclear Layer* (ONL) of retinal cells; while the non-classical opsins, such as melanopsin, are located in the *Ganglion Cell Layer* (GCL) & *Inner Nuclear Layer* (INL) (Figure 2.1, 2.3).

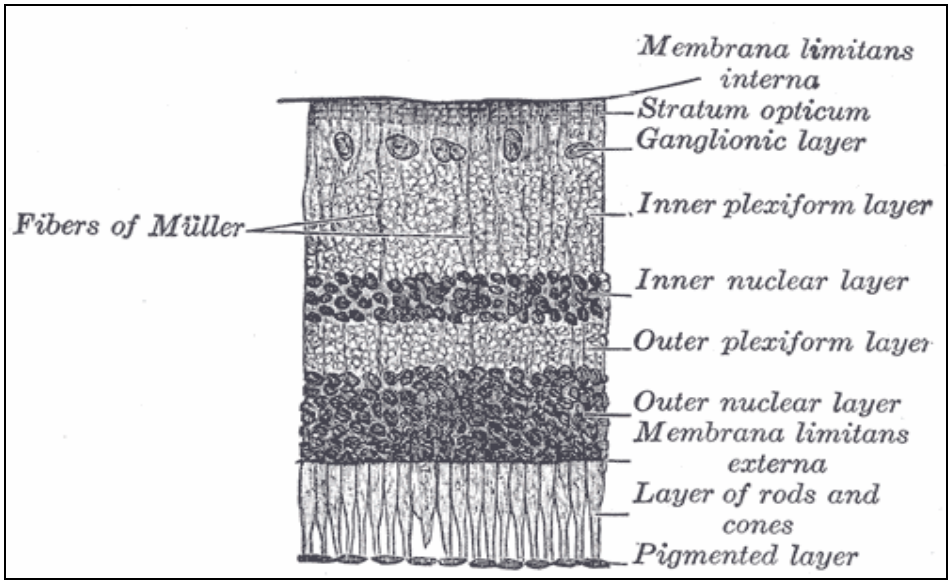


Figure 2.3: Layers of the retina

ONL is made up of nuclei and cell bodies of rod and cone cells; those cell types are sensitive to light by their classical opsin family proteins. Rod cells comprise of *rhodopsin* that plays role in perception of light, night vision. Cone opsins are responsible of color vision.

Melanopsin:

Melanopsin is a non-classical opsin protein expressed in GCL of retinal cells. This protein was discovered in the dermal melanophores of Xenopus [14]. Although this protein is present in the vertebrate retinal cells, it is evolutionarily similar to the invertebrate organism opsins; it has no resemblance to other vertebrate opsins in cellular localization and gene structure. Analogous to invertebrate opsins, melanopsin is assumed to make signal transduction through Gq family protein. Consequently, melanopsin should be distinct from other vertebrate opsins that play role in the visual roles; they are considered to play role in non-visual tasks, such as pupillary reflex and circadian rhythms [15].

The light response of GCL cells containing melanopsin were firstly verified by Berson et al [16, 52]. In those experiments the isolated cells were shown to be intrinsically photosensitive. Due to its intrinsic photo-action, melanopsin is assigned as a new photoreceptor class. Melanopsin was also shown to exhibit light and dark adaptation.

The expression of human melanopsin in mice paraneuronal cell line Neuro-2a, cells normally with no melanopsin production and no stimuli to light, made those cells reactant to light [23].

Another line of evidence that melanopsin is the photoreceptor of circadian rhythms is that its action spectra profile is similar to the action spectra of circadian responses [16, 17]. Melanopsin containing cells has a connection with SCN; the *retinohypothalamic track* provides projection of signal from GCL to SCN. The projection is via the neuropeptide PACAP (piturity adenylate cyclase activating peptide) of the ganglion cells [18, 19].

2.1.2 Clockwork Component

The clockwork component of circadian rhythms is mediated by the suprachiasmatic nuclei (SCN) located in the hypothalamus (Figure 2.1). The SCN contains several cell types, several different peptides and neurotransmitters. The clock works via the signal-pathway executed by photoreceptors of the retinal cells. The retinohypothalamic tract makes the connection between retinal ganglion cells and the SCN [1]. The SCN is also called as the master-clock of the body, since other pacemakers are found in other organs, such as liver. SCN is responsive to light, whereas liver is responsive to feeding for regulation of circadian rhythms. The cultured SCN is able to maintain its own rhythm without any need of external cues [68].

SCN regulates the circadian rhythm by gene expression and transcriptional factors. There are 2 proteins named; Clock (circadian locomotor output cycles kaput) and Bmal (brain and muscle aryl hydrocarbon receptor nuclear translocator) that behave as transcription regulators. Clock and Bmal proteins form a heterodimer and positively regulate the transcription of Per (Period) and Cry genes. Per and Cry make a complex and control their own expression by acting as negative regulators. This system is present in all cell types of mammals; the SCN is the center that regulates the clocks in all tissue types [1].

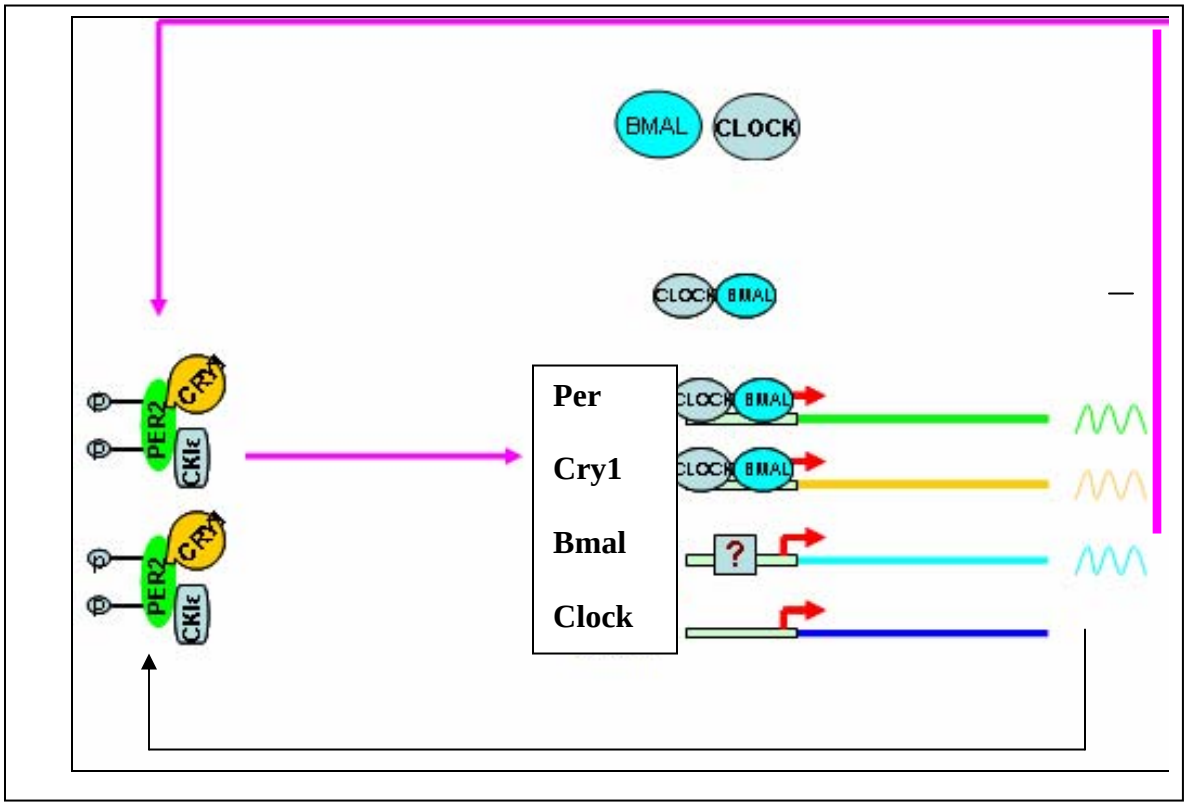


Figure 2.4: The regulation of circadian clock at SCN via the positive and negative transcriptional factors [1].

2.1.3 Output Component

The output component of the circadian rhythm is the changes in biochemical, physiological and behavioral functions of the organism. The clock at SCN leads to the secretion of prokineticin and transforming growth factor; those factors lead to the stimulation of other brain regions for alterations in the functions of the organisms [1].

The mental alertness, neuroendocrine levels, sleep-wake cycles, physical strength, blood pressure and viscosity, body temperature regulation, reproduction and many other functions of the mammals is adjusted via the circadian clock [54].

The clock is regulated via light-dark cycles; and the liver clock is also controlled via feeding habits as stated before. There are some other factors influencing the clocks in addition to the light-dark feeding cycles, such as; temperature and emotional stress. Those factors affecting circadian clock is named as “Zeitgeber” (meaning “time-giver” in German) [1].

2.2 Computational Studies Overview

Computational biology can be defined as the way of studying/solving a biological problem with use of computers. The bioinformatics tools, simulations, data analysis are used in order to find a solution to a specific problem. By the help of computational and experimental studies; advancement of biological knowledge is targeted.

Computational biology studies make use of bioinformatics tools in order to study proteins. There are various tools to model the secondary or tertiary structures of proteins; revealing the active sites of proteins; discovering the protein-protein interactions.

2.2.1 Protein Structure Modeling

Proteins have 4 types of structure;

1. Primary structure; the amino acid sequence of the protein
2. Secondary structure; local interactions of amino acids that are stabilized by hydrogen-bonding
3. Tertiary structure; overall structure of a protein in space; its 3-dimensional (3D) structure
4. Quaternary structure; the structure gained via the interaction of 2 or more proteins

The primary structure determines the secondary and tertiary structure of a protein. Especially, the tertiary structure is very important, since the 3D structure of a protein is crucial in the determination of its function.

The tertiary structure can be determined by experimental means; via X-ray, NMR techniques. Generally the experiments take a long time to set-up. When the determination of protein structure is not easy (i.e. the membrane proteins) or when the 3D structure has not yet been determined but it is needed in a very short time, the experimental method is not the sufficient solution. In such a case, the computational biology tools are very helpful; those tools make use of *comparative* and *ab initio modeling* in order to give the 3D model of the protein of interest.

Comparative modeling is a very efficient way of modeling proteins. It makes use of a database of proteins whose 3D structure is known. Those tools acquire the primary sequence of the protein of interest; aligns it to the sequences of proteins in its database. If a homology (similarity of sequences) is more than a pre-determined cut-off value, it uses those proteins in the database as template. Afterwards it models the protein of interest according to the template. The main idea behind comparative modeling is that similar sequences serve to similar protein functions; and similar functions should be directed by similar shaped proteins. This model is very trustable, but it only works when only related template structures are available. In this study, the reliable comparative modeling tools were determined and used for analysis. The evaluation methods for those tools is described in Chapter 3.

Ab initio modeling makes use of a database with small protein regions with known secondary or tertiary structures. When a query sequence is given to those tools, it determines the structure for small portions of this protein by alignment of query and database sequences; afterwards it joins those small fragments in order to get the full-view of the protein.

2.2.2 Protein-Protein Interactions

Protein-protein interaction can be defined as the association of 2 or more proteins. The interaction of proteins is essential for almost all processes in a cell of an organism; the signaling cascades are only possible via the protein-protein interactions. Those types of interactions can be studied via experimental ways by means of biochemical and molecular methods.

The interaction of proteins is possible via; electrostatic forces, van der waals forces, hydrophobic effects and hydrogen bonds. The major force that drives proteins to interact is the hydrophobic effect; other forces make the interaction stronger.

Protein-protein interactions characterized according to their [55];

1. Size and shape
2. Complementarity between surfaces
3. Residue interface propensities
4. Hydrophobicity
5. Cellular localization
6. Secondary structure
7. Conformational changes upon complex formation.

The protein-protein interaction prediction is possible with experimental techniques such as; yeast two-hybrid system, microarrays, affinity purification, and mass spectrometry. The false-positive results or results that are noisy can be detected by use of bioinformatics tools, such as Bayesian network models.

The interaction can also be deciphered via various bioinformatics approaches. Since the proteins are evolved, it is a very convenient way to use phylogenetic information to predict the interactions. In order to achieve a phylogenetic comparison; the sequence of the proteins of interest are used. Multiple sequence alignments are done by various tools, such

as BLAST [28]. The phylogenetic distance matrices are derived from sequence alignments. The interaction is determined to be possible if the calculated score of matrices is above a cut-off value.

Another way to predict protein interactions is possible with tertiary structures of known interacting proteins. The interaction face of known proteins is used to form a library for those tools. Sequence alignment is done to the interfaces of complexes; and residues that have a high frequency for a specific position are found. Those highly frequent regions are called hot-spots. This database containing known hot-spots and interface residues is then used to estimate the possible interaction between 2 proteins, with known 3D structure.

A useful method that searches protein complex formation is protein-protein docking. The tertiary structures of proteins of interest should be known for this method to be useful. This technique makes use of the geometric and physical properties of proteins. Method employs algorithms to calculate the interaction energy and/or shape complementarities of proteins. The calculation with the highest shape complementarity and lowest docking energy is chosen as the possible protein complex that was searched for.

The computational methods can give a good estimation on whether proteins interact or not; the interacting amino acids; the interaction energy. With usage of computational methods, the experimental procedures that would be employed are optimized and also time consuming steps are eliminated.

Chapter 3

BIOINFORMATICS TOOLS OVERVIEW AND EVALUATION

3.1 Bioinformatics Tools Overview

The analysis of interaction of proteins requires the tertiary structures of the proteins of interest as implied in Chapter 2. The functions of proteins, which are determined by their signal peptides and post-translational modification sites, are also very crucial determinants of protein interactions. The tertiary structure of Melanopsin is known [24]; but the structure of hCry2 has not been dechiphered by experimental or computational means. The experimental design for hCry2 structure determination is time-consuming; and it is hard to purify protein in its functional state due to its co-factors, as mentioned in the preceeding chapter. Only the experimental models of *A. thaliana* Cry (PDB id: 1U3C) and *Synechocystis* Cry (PDB id: 1NP7) proteins are known; those models have the photolyase similar domains, but the extended C-tail. Consequently, the computational structure modeling technique was chosen as a method to uncover hCry2 tertiary structure. The comparative modeling tools will be useful, since 2 sequentially similar Cry proteins have experimental models. Also *ab initio* modeling tools will be needed, since the C-tails of no cryptochromes have so far been experimentally modeled. In order to be able to make tertiary structure modeling; a reliable and efficient tools are needed.

A handicap of bioinformatics study is that there are numerous tools, which serve the same purpose. There are various free-ware bioinformatics tools on the internet, which can simply

be used on-line or by downloading the necessary files to your PC. Out of various tools it is very hard and time-consuming to find the most reliable one. Some assesments are held periodically, in order to find the most effective tool (i.e. CAPRI [56] for protein interaction prediction).

As in case of other prediction tools, there are numerous tools for tertiary structure prediction. Most of those tools employ the homology modeling technique; whereas a small number of them use the *ab initio* modeling techniques.

Critical Assessment of Techniques for Protein Structure Prediction (CASP) [57] is held biannually in order to find the most reliable tool for structure modeling. The result of CASP-2004 experiment was used to have an idea about the most reliable *ab initio* tertiary structure modeler. *Ab initio* modeling of hCry2 was decided to be done with Robetta tool of Baker's lab [36], the details of the tool are given in the Section 3.1.1. The Robetta tool makes a prediction in a long period of time, consequently a test for this tool could not be accomplished. Since the prediction of hCry2 structure by Robetta server took long time, comparative modeling tools, which were able to make prediction in hours, were also employed to have preliminary data for hCry2-Melanopsin interaction. A test was set-up in order to determine the most reliable homology modeling tool/tools. The details of this test and evaluation of the tools is given in Section 3.2.1. The evaluation of comparative modeling tools implied that; EsyPred [34] and Consensus [35] tools are reliable. Those two tools were used to model hCry2 protein by comparative modeling techniques.

As mentioned above, the functions of proteins are very important determinants of their structure. To be able to uncover the functions of hCry2 and Melanopsin; among various free-ware tools the reliable ones were chosen via literature search. The details of the tools is described in Section 3.1.2.

The final step of the tool evaluation was done for protein interaction tools. A literature search, results of CAPRI [56] assessment and several test were used to choose the tools. The description of tools is given in Section 3.1.3; the test set-up is given in Section 3.2.2.

3.1.1 Overview of Tertiary Structure Prediction Tools

Robetta:

Robetta is the name of a server that is designed for protein structure prediction [36]. The primary sequence of protein of interest is sufficient for structure prediction task. Server uses a method called *Ginzu* [58] in order to predict the structure.

According to the Ginzu method (Figure 3.1) the sequence is first scanned with BLAST [28], PSI-BLAST [59], FFAS03 [60, 61] and 3D-Jury [62, 63] algorithms respectively; BLAST is the most reliable tool, where 3D-Jury is the least reliable among 4 algorithms. The alignment with BLAST is done first; if there is a region with no homologous hits, then a match to that region is searched by other tools. Those regions with homologous sequences are modeled by *comparative modeling* technique. The sequence is partitioned into regions smaller than 17 amino acids length; they are modeled according to the sequence similarity that they share with their homologs found by 4 algorithms. Those modeled domains of protein are combined using an energy function [36, 58]; the continuity of the backbone is ensured by this function.

If there are regions with no homologs, those regions are passed to next step of Ginzu method. The uncovered regions are named as *linkers* if they are made up of less than 50 amino acids; otherwise the region is scanned via Pfam-A [64] and via Multiple Sequence Alignment (MSA) techniques. The domains found by those methods are modeled *de novo*/ *ab initio*.

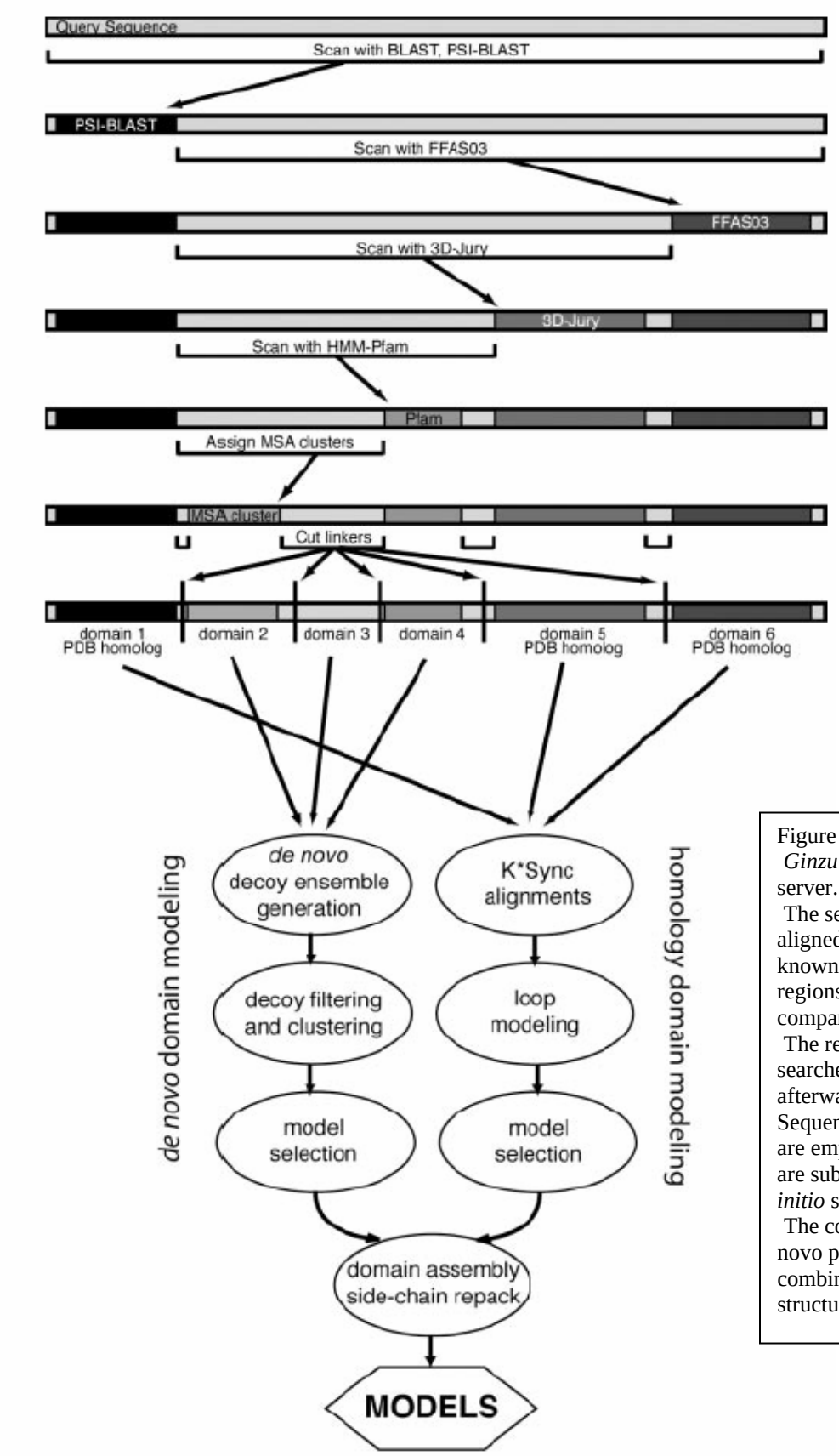


Figure 3.1:
Ginzu method of Robetta server.
The sequence is firstly aligned to databases of known structures. Those regions are modeled by comparative modeling.
The remaining regions are searched against Pfam-A; afterwards Multiple Sequence Alignments (MSA) are employed. Those regions are subjected to de novo/*ab initio* structure modeling.
The comparative and de novo predictions are combined to have the overall structure as a final step [36].

Robetta contains three- and nine-residue fragment libraries [36]. Those libraries contain the local folds found in the Protein Databank (PDB) [31]. The sequence that is going to be modeled *de novo* is partitioned into nine-residue fragments; the 3D folds from library are selected according to the proteins primary sequence. The nine-residue models are assembled through a scoring function and the overall model is completed. A Bayesian scoring function is used in Robetta [36, 58].

The server generates 10,000 decoys for the sequence of the protein of interest; 5000 decoys for the homologs; among those decoys 2000 and 1000 are chosen and assembled according to scoring function. Numerous filters are used to find the most reliable 10 models.

Comparative Modeling:

Comparative modeling can be defined as usage of a template protein, whose structure is solved, in order to model the structure of another protein. The main idea of this modeling technique is that similar primary sequences give rise to similar folds.

The number of protein folds in nature is comparatively less than the number of proteins. The number of structural motifs of proteins are believed to be approximately 2000. Different functioning protein structures are solved experimentally and deposited to PDB; consequently those proteins with different folds can be used as template in order to dechipher unknown protein structures.

Comparative modeling can be categorized as;

1. Homology Modeling
2. Protein Threading.

Homology modeling uses the idea that similar sequences result in similar folds. The method makes use of sequence alignment as stated in Section 2.2.1.

Protein threading makes sequence alignment of an unknown structure against a database of known structures, such as PDB. If there is any sequence similarity between the protein

of interest and database, the compatibility of sequence in the template 3D structure is analysed by a scoring function. For example, a positively-charged residue of template protein can be replaced by negatively-charged residue; or a small reside can be replaced by a bulky residue in the query sequence; those type of differences give rise to alterations in the tertiary structre of the protein. The scoring function yields the most appropriate model for query protein sequence.

Model Refinement:

The *ab initio* and comparative modeling tools both try to predict the structure of the backbone of protein accurately; but even the most reliable tools often predict the orientation of the amino acid side-chains inaccurately. The side-chains, which are mis-predicted, can occupy the same place in space and give rise to steric clashes. The steric clashes and wrong orientation prediction decreases the quality and reliability of the structure.

There are various tools to employ algorithms which analyze the orientation of side-chains to overcome structural problems. A reliable model refinement server is *What-If*. This server uses a program named WHAT-IF [37]; the program has approximately 2000 options such as; homology modeling, energy calculations, structure validation. In this study, this program was used to remove steric clashes from the structure and optimize the overall energy of the model.

3.1.2 Overview of Function Prediction Tools

The function of proteins is essential in determining the structure of protein –the folds- and the interaction with other proteins. The number of proteins with unknown functions are decreasing due to the improvements in proteomics, bioinformatics and experimental studies. The determination of protein function experimentally is takes long time to perform; consequently, the automated function prediction techniques are becoming more and more

preferable. As the function prediction is being a hot topic; the research done in this area is expanding.

In this study, the functions of proteins was tried to be uncovered by computation; via the function prediction tools. Literature search and tests were done to identify the most efficient tools.

The protein function is a broad area; for that reason it is categorized into different levels. A set of keywords were defined to study protein function;

1. Family; evolutionarily related protein group.
2. Domain; a part of protein that has the ability of self-stabilizing and folding independently.
3. Pattern; A group of amino acids which is repeated in different proteins with the same function. For example; Alanine-Leucine-X-X-X-Valine-Glutamine is a pattern, where “X” denotes any kind of amino acid. The patterns are usually short sequences that determine the function of proteins.
4. Motif; a distinct pattern of sequence that is conserved evolutionarily.

In this study the InterProScan [25], MotifScan [26], and ScanProsite [27] tools were used to identify possible signal peptides and post-translational modification sites of hCry2 and Melanopsin.

Proteins with defined functions are deposited under various databases, which later are going to be used to identify relationships in novel sequences, protein classification and function determination. Computational tools employ sequence alignment/ scoring matrices/ hidden Markov models and detect the known patterns and/or domains in a given sequence; consequently, the function of protein can be inferred.

There exist various databases and numerous tools making use of different databases. These databases are most effective when used together, rather than in isolation [25]; so most of the tools employ more than one database.

The tools are based on similarity-search of a given sequence in a given databes. The search is achieved via multiple alignments; those alignments can be global (making use of the whole protein) or local (making use of small portions of the protein). The matching residues are given a score. If the score is greater than a cut-off value the match is reported. The scoring can be done via profiles or hidden Markov models.

InterProScan tool makes use of InterPro database that integrates major protein databases, such as; PROSITE [70], PRINTS [53], ProDom [51], Pfam [49], SMART [41], TIGRFAMs [22], PIRSF [21], SUPERFAMILY [69]. PROSITE [70] is the database of profiles; PRINTS [53] is the database for unweighted motifs; other databases make use of hidden Markov models. The employment of various databases leads to a reliable and efficient results.

The properties of proteins from different databases are manually supplied to InterPro. The characteristic protperties for the same protein family or domain are grouped together into a single InterPro entry [25]; they have a single accession number. The 13.0 release of InterPro has 13147 entries; 3760 domains, 9080 families, 232 repeats, 32 active sites, 22 binding sites and 21 post-translational modification sites The database is available under Appendix A-2 in order to function prediction of any given sequence.

The ScanProsite tool [27] uses PROSITE database. The tools make use of multiple sequence alignments and weight the positions of residues; it searches for the occurence of patterns, profiles and motifs in the database.

MotifScan tool [26] makes use of MyHits databases. This database contains various other databases; PROSITE [70], Pfam [49], InterPro [25] ,TIGRFAMs [22], and HAMAP [67]. The tool tries to dechipher all of the motifs in a given sequence.

3.1.3 Overview of Protein-Protein Interaction Analysis Tools

Protein-protein interactions are determinants of the cellular functions. The biological processes are available only if the protein interactions are guaranteed to occur properly. In order to have a better understanding signal-pathways and biological processes, the protein interactions should be discovered.

Investigation on protein-protein interactions can be achieved by experimental or computational approaches. Yeast two-hybrid essays, mass spectrometry and other experimental approaches enable us to study protein-protein interactions. Another important improvement is the development in computational analysis tools. The analysis tools are being developed and updated in order to study the interactions in atomic level; and the tools also shortens the research time. The computational analysis of interactions try to reveal whether a group of proteins interact or not; the strength of the interaction; the possible conformational changes of the interacting proteins. The computational study of protein interactions is named as “protein-protein docking”. The term can be defined as the study of molecular structure determination of interacting proteins. The computational interaction studies can be achieved by various different methods.

The prediction of protein interactions is generally based on the shape complementarity of the proteins and minimization of total energy of protein complex system. The shape complementarity basically makes use of the 3D shape of protein surfaces and determines whether those surfaces fit or not. The energy of complex is minimized by a function containing the information about the thermodynamics and kinetics of the system. The electrostatic terms, van der Waals energy, hydrophobic-hydrophilic properties of the residues are some of the variables used by energy minimization functions. The shape complementarity, electrostatic forces, or any other properties are applied to protein complexes; a simulation is carried by computers to decide whether the proteins interact or not.

The interaction prediction is based on either rigid or flexible docking. In the rigid docking the proteins' bond-angles, bond-lengths and torsional-angles are fixed; no modification on protein structure occurs during the docking process. The ridig docking is very helpful in determining whether proteins form complex or not. In case of *in vitro* conformational changes this method is not very useful; since the structural alterations cannot be mimiced by the method; consequently the results may be inefficient. In flexible docking some, but not all, conformational changes are permitted. In other word flexible docking is much sufficient and reliable than rigid docking; but when computer time use is taken into account rigid docking is preferable.

The docking can be achieved by various search methods as stated before. Some of the docking tools employ Fourier Transform correlations; they make use of *Fourier transform functions* as a base. The inputs such of electrostatic potentials are represented as expansions of Fourier basis functions. There are other tools making use of *reciprocal space methods* [64]. The method represents the proteins as cubic lattice; afterwards translation of one lattice is achieved with the use of a lattice vector to the other lattices. Simultaneous translations are made via convolution theorem application, until sufficient interaction is determined. The steric and electrochemical properties of the proteins can be used in the scoring functions of reciprocal methods. Another widely used method for docking analysis is Monte-Carlo. Monte-Carlo methods make random-steps over an initial conformation; each step is accepted or rejected with use of a scoring function. The method is very efficient, but some good protein-complex configurations can be missed since the movements are random.

Each docking analysis tool has a selection criteria among the possible complex structures found by search methods. The selection can be achieved via [65];

1. Measurement of shape complementarity
2. Free energy measurements

3. Phylogenetic correspondance of the interaction region
4. Heuristic scores based on residue contacts.

To conclude; the possible protein interactions can be studied via a tool based on rigid or flexible docking; a type of search method and a selection criteria. In this study 2 tools were chosen for protein interaction analysis; Hex 4.5 and AutoDock.

Hex:

Hex 4.5 [43] tool was developed by David Ritchie, at University of Aberdeen. The tool is used for both protein docking and molecular superposition. This program uses polar Fourier correlations to increase the calculation speed. It makes use of electrostatic force and shape complementarity properties of the protein complexes.

Most of the prediction tools use the rigid docking technique, and reduce the complexity of the problem. There is a withdrawal of such tools, the methods they employ need a long period of time after the calculations in order to remove the steric clashes. Hex tool overcomes this problem by its Fourier correlation approach, since this method makes use of a Cartesian grid that removes steric clash possibilty.

The Fourier approach has other advantages such as; the calculation of van der Waals and electrostatic force field models is accelerated; hydrophobicity of proteins can be taken into account; low-resolution docking can be achieved [43]. There are some disadvantages of the tooll; requirement of a large grid and calculation of each rotational change increases the time and memory needs.

The Hex tool can be freely downloaded (Appendix A-8).

The tool is very efficient if a preliminary information about the protein complex is known; it gives still reliable results for the protein complexes that does not have any pre-information. The tool accepts .pdb format protein or DNA files.

AutoDock:

AutoDock is an automatic docking method for prediction of macromolecule-ligand complexes [38]. The tool is also capable of calculating the free-energy of the system. The tool has three search method choice; Lamarckian genetic algorithm, Monte-Carlo simulated annealing and traditional genetic algorithm. Lamarckian genetics model is the basis of the tool in which environmental adaptations of an individual's phenotype are reverse transcribed into its genotype and become heritable traits [38]. The Lamarckian genetic algorithm is found to be the most succesful method among three by the programmers of the tool.

The tool uses a flexible ligand and a rigid protein. The ligand can be an inorganic or organic chemical or polypeptide. The protein is put into a three-dimensional grid. The dimension of the grid, the rotatable bonds of the ligand, the start region of the ligand, search method is specified by the user.

The energy calculation of protein and ligand at each point of the grid is achieved by the following method;

1. A probe atom is placed at each grid point
2. The energy between protein and the probes are calculated & assigned to each grid point
3. Affinity grid for the atoms of ligand and charge on each probe is calculated
4. The eight grid points around each atom of the ligand are used to calculate the energy field of the ligand.

Those calculated energies of ligand are stored in a file to be used in the docking process.

The docking simulation makes use of rigid protein and flexible ligand; the ligand makes random walk aroun the protein. By each step, the ligand changes its orientation; and the energy of the ligand in its new region is easily found from the pre-calculated and stored

energy values. The new energy of the ligand is then compared with its energy at its previous orientation; if the new energy is lower, then the previous energy is rejected. The simulation continues until a minimum energy is reached.

AutoDock is a very reliable tool, which has a high score in CAPRI[56]. It is better than Hex in the sense that it allow ligand to be flexible, while Hex makes rigid docking.

3.2 Bioinformatics Tools Evaluation

The numerous bioinformatics tools help researchers to target problems, but on the other hand the reliability and efficieny of those tools are still problematic. As stated before assessments such as CASP [57], CAPRI [56] help us to discriminate the best tools among good ones. Unfortunately, most of the tools in those assessments are very expensive, although some of them are free-wares. Consequently, the good tools that are free, but not employed in those assessments should be revealed. In this study, a free-ware and good (according to CASP [57]) *ab initio* modeler was used; Robetta server [36]. The tools that make use of comparative modeling were evaluated as described in Section 3.2.1.

The AutoDock 3.0 and Hex 3.1 were both evaluated and ranked good in CAPRI [56]; but several tests were used for Hex 4.5 tool, since this edition of the tool had not been evaluated.

3.2.1 Evaluation of Comparative Tertiary Structure Prediction Tools

In order to find a good tool for comparative protein structure prediction, a *library* of the tools was created (refer Supplementary # 1). The test for the most reliable tool was decided to be done with *A. thaliana* Cry protein. As described in Chapter 2, the cryptochromes have a similar domain (DNA photolyase domain); consequently the usage of *A. thaliana* Cry as query would also let us to tool that is proper for hCry2.

The experimentally discovered structure and the amino acid sequence of *A. thaliana* Cry is obtained from Protein Data Bank [31]. The pdb code for this protein is 1U3D, the protein is made up of 483 residues. The amino acid sequence of *A. thaliana* Cry protein is given to all 3D prediction tools in our library as query.

The predicted 3D structures should be compared with the known 3D structure of *A. thaliana* Cry; to determine which homology modeling tool is the most reliable. In order to be able to compare two 3D structures, a reliable *superimposition tool* was needed. A library of free-ware superimposition tools was created (refer Supplementary # 2).

To determine the most reliable superimposition tool among them, the real structure of A. *thaliana* Cry is compared with itself. All of the tools was reliable, but the ones giving the most detailed results were; *Combinatorial Extension (CE) Method* [32] and *FatCat* [33]. These tools also give results rapidly.

The predicted and real structures of *A. thaliana* Cry were compared by CE method and FatCat (Table 3.1, refer Supplementary # 3 for all of the results).

The protein is 483 residues long, as stated before. Although the tools make homology search through databases and have the ability to find the protein itself; most of the tools could not be able to model the protein fully. Most of the tools found more than zero root mean square deviation (rmsd); but the rmsd levels were in tolerable range. Some of the tools even did not give any result.

The most reliable 3D prediction tool is found as *EsyPred* [34] and *Consensus* [35]. EsyPred was able to model the protein fully; due to its algorithm, or due to its updated database. Consensus was able to model the protein with no deviation from the real structure; most probably due to its algorithm, since it does not have real structure protein in its database, but it is able to configure the energy of the protein accurately.

Consequently, hCry2 protein homology prediction is made with Consensus and EsyPred tools.

Table 3.1: The superimposition of real *A. thaliana* Cry structure with several predicted models are given. The protein length was 483 residues; EsyPred is the most efficient modeller since it could model the whole protein. Consensus is also efficient, since the rmsd value of the model is zero.

Program Name	**EsyPred Model**
rmsd	0,2
Z-Score	8,3
Alignment	483
Gap	0
Program Name	**Consensus Model**
rmsd	0
Z-Score	8,2
Alignment	388
Gap	0

3.2.2 Evaluation of Protein-Protein Interaction Analysis Tools

After the modeling of hCry2, the interaction of this protein with Melanopsin was studied. As stated before AutoDock 3.0 and Hex 4.5 tools were chosen to study the protein docking problem. Both of these proteins are evaluated as good (AutoDock is better) in CAPRI [56]. The Hex 3.1 was evaluated in the assessment; but the up-dated version “4.5” was not. The up-dated version is expected to be better, but since the tool is new it can be problematic sometimes due to a bug, etc.

Hex 4.5 tool was used to evaluate pairs of proteins, whose interaction properties are known. The used pairs was; Per2-Bmal, Clock-Bmal, Tpr- Hsp, Hemoglobin-Heme, Cry-Fad, Cry-Mtfh, *E.coli* Photolyase-Fad, *E. Coli* Photolyase-MTHF. All of those interactions were predicted accurately by Hex 4.5 tool (Table 3.2).

The shape complementarities and electrostatic forces between melanopsin and hCry2 had been investigated by Hex 4.5; afterwards dock between these proteins investigated by the 3.0.5 version of AutoDock tool.

Table 3.2: The docking results of various protein pairs by Hex 4.5. The expected and Hex results are the same; revealing the reliability of the tool once again.

Protein 1	**Protein 2**	**Docked ?/Expected**
Bmal	**Per / Clock**	Yes / Yes
Hemoglobin	**Heme**	Yes / Yes
Cry	**Fad / Mtfh**	Yes / Yes
***E. coli* photolyase**	**Fad / Mtfh**	Yes / Yes
Tpr	**Hsp**	Yes / Yes

Chapter 4

COMPUTATIONAL METHODS

4.1 Methods for hCry2 Protein

The primary sequence of hCry2 is obtained from NCBI (accession number: NP_066940). The sequence was used as input in all of the methods described below.

4.1.1 Cellular Localization, Post-translational Modification Site and Signal Peptide Site Analysis

The probable signal peptide and post-translational modification sites of hCry2 are searched via *InterProScan* [25], *MotifScan* [26] and *ScanProsite* [27] tools. The amino acid sequence of hCry2 is given as query to the servers of these tools; and the results were taken either interactively or via e-mail.

As controls, the active sites of *E. coli* DNA photolyase, *A. thaliana* Cry, Mouse Cry2 were discovered by the same tools.

The molecular function, cellular component and biological processes of the hCry2 were determined by *JaFa* [46] tool.The cellular localization of the protein was also examined by *PSORT II* [47] tool.

4.1.2 Secondary Structure Analysis

Three reliable secondary structure prediction tools used for analysis; *Rosetta-HMMER* [29], *Rosetta* of Baker's lab [30] and *PredictProtein* [48]. Rosetta-HMMER and PredictProtein give the results within a couple of hours, via e-mail. Rosetta makes prediction in a long period –up to months-; registration through the server is necessary to be able to make runs. The primary sequence of hCry2 is given as query to the tools.

4.1.3 Tertiary Structure Analysis

The *Robetta* [36] tool was used for *ab initio* modeling; while *EsyPred* [34] and *Consensus* [35] tools were selected for comparative modeling.

The results from Consensus and EsyPred were obtained in a couple of hours via e-mail, while the 10 results from Robetta obtained after 5 months from the submission date.

Refinement of the Tertiary Structure:

The Robetta model with the most structural stability is determined by WhatIf tool [37]. The secondary structure predictions and the secondary structures of the most stable model were compared. The changes in the model were done according to those comparisons.

After the appropriate modifications; the most stable model was refined by the "remove bumps" and "complete a structure" options of WhatIf tool.

The 3D similarity search of the hCry2 model:

EMBL-Dali [42] was used to find structures that have resemblance to hCry2 in tertiary structure. This tool makes 3D superimposition of a given protein (.pdb file) with the proteins at PDB. The results of this tool deciphered new domains on the human Cry2.

The superimposition of 3D structures of hCry2 and found homologs is made by *FatCat* [33] and *CE* [32] tools.

4.2 Methods for Melanopsin Protein

The tertiary sturcture of hamster Melanopsin was determined computationally by *Hermanna et al* [24], as stated in Chapter 3. The similarity of human and hamster Melanopsin sequences was compared by BLAST [28] algorithm. The resemblance between two proteins was high in percentage (83 %); consequently the 3D model of hamster protein was used.

Since the Melanopsin is a transmembrane protein, the hCry2 cannot reach and dock to its membrane parts; consequently the membrane, cytosolic and extracellular parts of this protein should have been uncovered. This search was done by the SPLIT [40] tool. The primary sequence of Melanopsin was given as query, and the result was taken interactively.

The similarity of cytosolic parts of the hamster and human Melanopsin is very crucial; since the possible interaction between melanopsin and cryptochrome may be via the cytosolic parts of Melanopsin. The sequence similarity between cytosolic loops of hamster and human Melanopsin were investigated by BLAST. The different residues were determined; those different residues of hamster Melanopsin were altered by its corresponding residue in human Melanopsin via the “mutate a residue” option of the WhatIf tool. This option of the WhatIf tool makes one mutation per run; consequently numerous runs were done. The tertiary structure of hamster Melanopsin is given to the server (in .pdb format) as query; and the results in .pdb format were taken interactively.

The final mutated tertiary model of Melanopsin contained the cytosolic loop residues of human Melanopsin. This model was refined by “remove bumps” and “complete a structure” options of WhatIf tool.

To visualize the tertiary structures of both proteins; **WebLab ViewerPro Version 4.0** and **RasMol Viewer Version 2.7.2.1.1** tools were used.

4.3 Methods for Protein Docking

The electrostatic forces and shape complementarities between Melanopsin and hCry2 proteins were investigated by Hex 4.5; afterwards detailed docking analysis between these proteins investigated by the 3.0.5 version of AutoDock tool. By this way, the docking made by AutoDock would become much more precise and the number of the trials would decrease.

4.3.1 Hex Methods

The shape complementarities and electrostatic forces between melanopsin and hCry2 were investigated by Hex 4.5.

The control for Hex 4.5 reliability was achieved by making several docking runs with known interacting/non-interacting pairs of proteins as implied in Section 3.2.2. For further check of the reliability of Hex 4.5 tool and for control of docking between hCry2-Melanopsin, *E. coli* DNA photolyase and *A. thaliana* Cry were docked to Melanopsin.

Hex 4.5 tool has a preference to save the 100 most probable results of the docked pairs. Generally, but not always, the maximum 20; and more specifically maximum 5 are found as the most probable docked conformations by *Ritchie et al* [43]. In our study, the python code “energy.py” (refer Supplementary # 4) was written in order to be able to choose the energetically most probable dock conformation among these 100 dock results. The script calculates the electrostatic and Van der Waals energies for the atoms closer than 7 Angstroms.

Docking Parameters for Hex 4.5:

The molecules that were studied were very large, consequently we asked the advice of the creator of Hex 4.5 tool (Dr. David Ritchie), who recommended *the macro-sampling option* of the tool.

Harmonics of order l-5 were enabled by the “Harmonics” option of the tool; then by “Macro-Sampling” option, macro-sampling was committed. Afterwards the “Docking Control” options were enabled; search mode was chosen as “full rotation”, correlation type was chosen as “Shape + Electrostatics”, receptor range and ligand range was chosen as “45”, all other options were remained as default.

Shape + Electrostatic Search for Docking:

The interaction between Cryptochrome and Melanopsin is possible via the three cytosolic loops and the C-terminus of Melanopsin. The Melanopsin was partitioned into its cytosolic loops and its C-terminus; the coordinates of each of those 4 protein portions were saved as .pdb files. The charges of the atoms were added by AutoDock tool. Afterwards possible interactions were searched via Hex 4.5 tool.

An interaction of cryptochrome protein was discovered by *Sang et al* [45], in that study the *A. thaliana* Cry was found to make homodimer. The N-terminal of *A. thaliana* Cry was discovered to make homodimer in light dependent manner, and this dimerization was found to be essential for C-terminal mediated signal transduction of Cry2. To determine if a similar homodimerization occurs in hCry2, the N-terminal of the protein was docked to itself.

The possible interactions between the N- and C-terminals of hCry2 was investigated.

In order to be able to determine whether the structurally similar domains of hCry2 (found by EMBL-DALI [42] tool) and hCRY2 have only structural similarity or also functional similarity; the domains were docked to the cytosolic part of the Melanopsin.

To summarize, the following docking searches were performed via Hex 4.5;

1. hCry2 → receptor; Casein kinase 1E → ligand (known to dock, used as control).
2. hCry2 → receptor; cytosolic part (all three loops and C-tail) of Melanopsin → ligand.
3. hCry2 → receptor; first cytosolic loop of Melanopsin → ligand.

4. hCry2 → receptor; second cytosolic loop of Melanopsin → ligand.
5. hCry2 → receptor; third cytosolic loop of Melanopsin → ligand.
6. hCry2 → receptor; C-tail of Melanopsin → ligand.
7. *A. thaliana* Cry → receptor; cytosolic part of Melanopsin → ligand (negative control)
8. *E. coli* DNA Photolyase → receptor; cytosolic part of Melanopsin → ligand (negative control).
9. hCry2 USP-similar domains → receptor; Melanopsin → ligand.
10. hCry2 N-terminal → receptor and ligand (homodimerization occurance?)
11. *A. thaliana* Cry N-terminal → receptor and ligand (known to dock, used as control for hCry2 homodimerization).

Since *A. thaliana* Cry2 is known to make a homodimer; its resulting dock energy taken from Hex 4.5 simulations is determined as cut-off value for Hex 4.5 tool.

4.3.2 AutoDock Methods

AutoDock is able to take one full-length 3D protein as receptor, and approximately 10 amino acid-length protein portion as ligand. A 10 amino acid sequence is generally optimum for a ligand of AutoDock, because this tool has a rotatable bond limit; it gives permission to maximum of 32 rotatable bonds.

Ligand preparation:

Melanopsin was portioned into subgroups of different lengths. Only polar hydrogens were added to the protein portions; afterwards Kollman charges were added by ADT (AutoDock Tool).

The portions smaller than 10 amino acids, have maximum 32 free torsional-bonds; so they are determined as *flexible* by AutoDock. The portions larger or equal to 10 amino acids

have minimum 32 torsional-bonds, so they were made *rigid* by the AUTOTORS utility of AutoDock. The ".pdbq" files were formed.

Receptor preparation:

Polar hydrogens were added to hCry2 and Kollman charges were added by ADT. The ".pdbqs" file was formed.

Grid prepration:

The grid map was determined by ADT. The receptor was taken as grid center. A grid of 126x126x126 with a spacing of 0.375 Angstrom between grid points was prepared.

The protein was not small enough to fit into one grid of 126x126x126 dimensions. Consequently, 5 different grids were formed. The center of each grid in x-y-z format is as follows;

1. **36, 3, 5.4**
2. **18, -8, 5**
3. **13, 33, 5.4**
4. **15, -10, 5**
5. **40, 3, -10**.

Dock preparation:

The efficient parameters for blind docking with AutoDock were discovered by *Hetenyi et al* [39]. Those efficient parameters are used for our docking purposes.

Lamarckian Genetic Algorithm was chosen as docking search parameter. For the flexible ligands population size was set to 250; while it was chosen 50 for rigid ligands. For flexible ligands 100 runs (change due to limiting time = 120 hours) were done; whereas 50 runs were made for rigid ligands. For all of the ligands, maximum number of energy evaluations

was set to 1 500 000; number of generations was set to 50 000. The mutation rate (0.02), the cross-over rate (0.8).

4.3.3 Determination of the Residues at the Interaction Surface

The HEX and AutoDock tools both gives results in *.pdb* format. Those *.pdb* files contain the information about the coordinates of hCry2 and Melanopsin proteins. The residues at the interaction surface can be analyzed by various tools. In this study the tool called *NACCESS* [66] was used to determine the residues at the interfaces. *NACCESS* calculates the accessible surface area of a protein complex.

NACCESS runs in Linux environment and its' codes are written by FORTRAN. The logic of *NACCESS* is given at Figure 4.1.

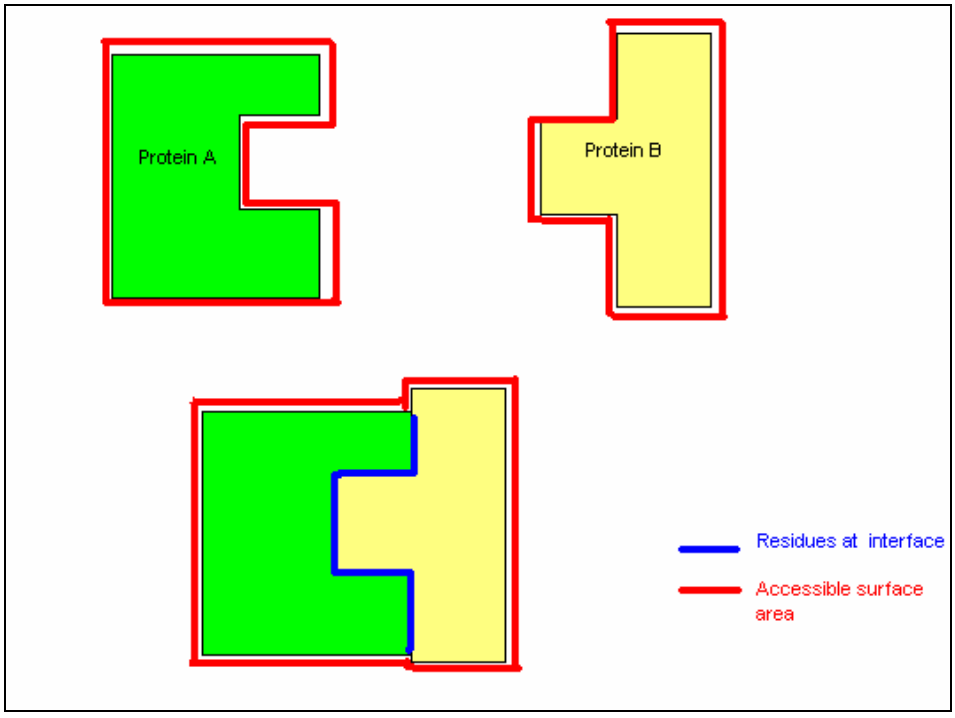


Figure 4.1: The two proteins Protein A and B are assumed to make a complex. Protein A is shown with green color; Protein B is shown with yellow color.
The solvent accessible surfaces of the two proteins and the protein complex is indicated with red line; the residues at the interaction surface are indicated ith blue line. As can be seen, the residues at the interface are no longer accessible to solvent when the complex is formed.

The tool needs the *.pdb* files of protein complex and the proteins of the complex alone as input. It simply calculates the accessible surface areas of proteins alone and the protein complex. It gives 3 output files; *.rsa*, *.log*, *.asa*. The *.rsa* file is important; it contains the absolute accessibility and relative accessibility informations.

The absolute accessibility values in the *.rsa* files are taken into consideration in this study. If one of the residues' absolute accessibility value in protein complex is decreased by >=1 Angstrom^2 when compared to its absolute accessibility value in protein alone, it means that this residue is at the interface of the complex.

Chapter 5

EXPERIMENTAL METHODS

The possible interactions between hCry2 and Melanospin proteins were determined via computational tools as described in Chapter 4. The results of those methods were analyzed and experimental procedures were decided to be employed in order to verify the results of the computational methods. Molecular biology, biochemistry and biophysics approaches were combined to substantiate the probable interaction.

In this study, firstly the DNA of the mouse Cry2 and Melanopsin proteins were constituted; those DNAs were either directly expressed in cell cultures or they were fused with fluorescent labels for further analyses. The expressed proteins were examined via microscopy, biochemistry or biophysics techniques.

5.1 Construction of Expression Vectors

DNA fragments with appropriate restriction sites were synthesized by PCR technique via using mouse Cry2 and mouse Melanopsin as template.

Amplified Cry2 DNA PCR products were purified and digested with XhoI. The pEGFP (Clontech Inc) were digested with same restriction enzyme and ligated with PCR product. pEGFP-mCRY2 were selected using alkaline based method miniprep. Amplified Melanopsin DNA was similarly isolated, digested with KpnI and EcoRI enzymes. Digested Melanopsin DNA was ligated with pDSred1.1 mammalian expression vector.

After transformation, the cells were incubated at 37°C shaker in LB medium for 30 minutes. The transformed pEGF plasmids were spread onto LB plates with kanamycin; and transformed p517 plasmids were spread onto LB plates containing ampicillin. All plates incubates over night at 37 ^{o}C incubater.

Colonies were picked by toothpick the day after spreading and grown overnight in LB containing appropriate antibiotic. Mini-prep was performed to extract the plasmid the next day (refer Supplementary # 5 for detailed Mini-prep protocol).

In order to be sure that we had the DNAs of interest; pEGF and pEGF-Cry2 samples were digested with EcoRI. The digestion should yield in DNA fragments of known size; consequently the digested samples were controlled by gel-electrophoresis techniques. The control for pDsRed1.1 and pDsRed1.1-Melanopsin was performed by BamHI digestion and gel-electrophoresis techniques. The samples containing DNAs of interest were determined. Maxi-prep was performed to extract plasmids in high amount (refer Supplementary # 6 for detailed Maxi-prep protocol).

The DNA was purified by agarose gel purification technique, in case it contains any contamination. The pure DNAs were stored at -20^{o}C.

The technique is represented schematically in Supplementary # 7.

5.2 Expression of Proteins

The DNAs that were purified and stored were used for protein porduction, in other words protein expression. The vectors carrying the DNA of interest should be transferred to a cell culture in order to able to synthesize the protein of interest from that DNA. 293 T cells were used in our experiments. 293 T is a derivative of 293 human renal epithelial cell line that is transformed by adenovirus E1A gene product. The cells in this study were prepared

by the transient transfection method according to $Ca_3 (PO_4)_2$ transfection protocol (refer to Supplementary # 13).

The cells that were taken from a definite line was stored in -80^oC. They were taken out of -80^oC and heated rapidly when they would be grown in cell culture. The cell culture used is DMEM supplemented with 10 % FCS (Fetal Calf Serum), 2mM Glutamine and 1000 IU Penicillin/Streptomycin.

For 10 cm dish, 10 ml culture medium was used. The cells were plated and when they reach approximately 70 % confluency the transfection was accomplished. On the day of transfection, 10 µg DNA was added to double-distilled H_2O in 15-ml sterile tube, then 155 µl 2M $CaCl_2$ was added. 1250 µl of 2×HEBS was added dropwise to the solution, while mixing the solution with a pipette by up-and-down technique. After 8 minutes, this solution was added evenly over the cells in culture. The medium changes it color from pink to orange. The culture then was incubated for 7-11 h. A dust like precipitate was observed. After incubation, the cells were rinsed by PBS once and fresh 10 ml medium was added. Cells were harvested 24-48 h after transfection. The proteins were produced as long as the cells are alive. For microscopy analysis, the cells were put into two cover glasses and sealed with epoxy. For biochemistry techniques, the cells were collected in PBS solution and stored at -80^oC.

5.3 Localization of Proteins

The fluorescence microscopy technique was employed to determine the locations of proteins that are tagged with fluorescent proteins. The flourescent microscopy at Cerrahpasa University, Biomedical Genetics department was used to visualize the proteins.

A collaborator group (Dr. Alper Kiraz) in our institute had used confocal microscopy for detailed analysis of protein localizations and the possible Melanopsin-Cry2 interaction. The confocal microscopy was used to scan the whole cell point by point.

488 nm Ar-Ne laser as an excitation source for donor fluorophore, GFP and 532 nm diode pumped Nd-YAG laser as an excitation source for acceptor fluorophore, RFP were used in Dr. Kiraz's laboratory [71]. FRET technique was employed in their study.

Chapter 6

RESULTS and DISCUSSION

In this chapter computational and experimental results are discussed.

6.1 Results and Discussion of Computational Methods

6.1.1 hCry2 Protein

Cellular Localization, Post-translational Modification Site and Signal Peptide Site Analysis:

The probable signal peptide and post-translational modification sites of hCry2 are searched via *InterProScan* [25], *MotifScan* [26] and *ScanProsite* [27] tools (Table 6.1 and 6.2).

As shown in Table 6.1 and Table 6.2 hCry2 have motifs which can be subjected to the post-translational modifications such as Lipoprotein Binding site; and Calpain Cat site.

The DNA photolyase was used as negative control, since there is no psot-translational modifications in *E. coli*. The *A. thaliana* Cry is a plant cryptochrome, consequently it plays role in a different signaling cascade [2, 20] and may have different functional sites. *A. thaliana* Cry was used as negative control. Mouse Cry2 is used as positive control, since it is a mammalian cryptochrome; since it needs post-translational modifications and it should have similar sites to hCry.

The *InterProScan* tool displayed 2 domains; *photolyase similar part* at the N-terminal and a *signal peptide part* at the C-terminal of hCry2 (Table 6.1). The N-terminal result was expected, since it is known that cryptochromes have similar structure to the DNA photolyase; while the C-terminal result was novel, since the C-terminal is believed to play role in signaling pathway, but there is no known evidence of it [1].

Table 6.1: The active sites and post-translational sites of hCry2 found by InterProScan tool. The numbers at the left column indicate the region of protein, at which the sites are found (the amino acid numbers at the primary sequence of hCry2).

Location	InterProScan
530-546	Sensor Histidine Kinase/Response Regulator (Autophosphorylation)
545-569	Stathmin Motifs (for signal transduction)
550-568	Putative metal-dependent phosphohydrolase HD domain-containing

The C-Terminal was divided into fragments of length 10, 25, 50 amino acids, in order to identify the exact location of the signal peptide. These fragments were run on the InterProScan. 4 signal peptide regions were identified at the C-terminal. These 4 regions were subjected to *BLAST* [28] to uncover their possible roles. The *530-546* amino acids part found to have “Sensor histidine kinase/response regulator from Caulobacter crescentus”; the *550-568* portion gave the result as “Putative metal-dependent phosphohydrolase HD domain-containing protein”; the *545-569* part determined as “Stathmin from mouse, chicken, human, etc”, while the fourth part did not match to any known domains.

The functions of these domains that were found at the C-terminal of hCry2 by *InterProScan* are;

1. *Stathmin* protein inhibits microtubule formation; its phosphorylation is required for its activity. The active site of stathmin is known to contain an alpha-helix; the

region that is approximated to sit on hCry2 is also predicted to have alpha-helix structure by Robetta model #3.

2. *Histidine/Kinase Response Regulator* is mostly seen in prokaryotic organisms, but known to exist in some eukaryotes, too. It has the ability to self-phosphorylation. This region may act for phosphorylation activity of hCry2. It is known that *A. thaliana* Cry undergoes blue-light induced autophosphorylation [42, 43]. The autophosphorylation reaction takes place via the FAD cofactor of Cry2. Since hCry2 has similar structure to *A. thaliana* Cry and also has FAD as cofactor, it is possible that hCry2 is able to make autophosphorylation.

The *MotifScan* gave many active sites as result (Table 6.2), some of which were expected and some were startling;

1. The *RGD*, the cell-membrane attachment site, is guessed by *MotifScan* to sit in Cry2 and Cry1 of human, Cry2 of Mouse; but not in *A. thaliana* Cry and *E. coli* DNA photolyase. The *ScanProsite* tool also predicts cell-attachment site sitting in hCry2.
2. *Calpains* play role in cytoskeletal modeling, cell-mobility, apoptosis/ necrosis [44]. The Calpain domain that estimated to sit in hCry2 protein is a very large region, with high homology value. These increase the probability of having Calpain region in hCry2. The Stathmin region and Calpain region found on hCry2 indicates that this protien may play a role in cell cycle.
3. The *Lipoprotein Attachment* site may act as a transporter, to make the hCry2 bound to the cell-membrane. The cell-membrane is made up of lipid-bilayer (hydrophobic molecules), and hCry2 should be encapsulated by hydrophobic molecules to have contact with this lipid bilayer.

The amidation and mystrilation results found by *MotifScan* and *ScanProsite* tools (Table 6.2) are generally predicted for even bacterial proteins; consequently they can be thought as false-positive results. For that reason the detailed information about them is not given here. Cryptochromes are known to interact with casein kinase 1ε [1]; so the results about this site found by *MotifScan* and *ScanProsite* tools are expected.

The cGMP-dependent phosphorylation site and GTP-binding motif predicted by both *MotifScan* and *ScanProsite* tools are prominent, since those motifs are very crucial in signal cascade pathways of mammalians. Those motifs can be thought as the evidences that prove hCry2 may play role in a G-protein coupled receptor (GPCR) signaling pathway; the protein may have the ability to bind GTP, and elongate the signal from GPCR (Melanopsin) throughout the cell. Consequently, those predicted sites support and broadens our hypothesis.

The novel funtional sites found by computational tools are shown on the 3D model of hCry2 in Figure 6.1. The modeling and refinement of this 3D structure will be discussed in this section.

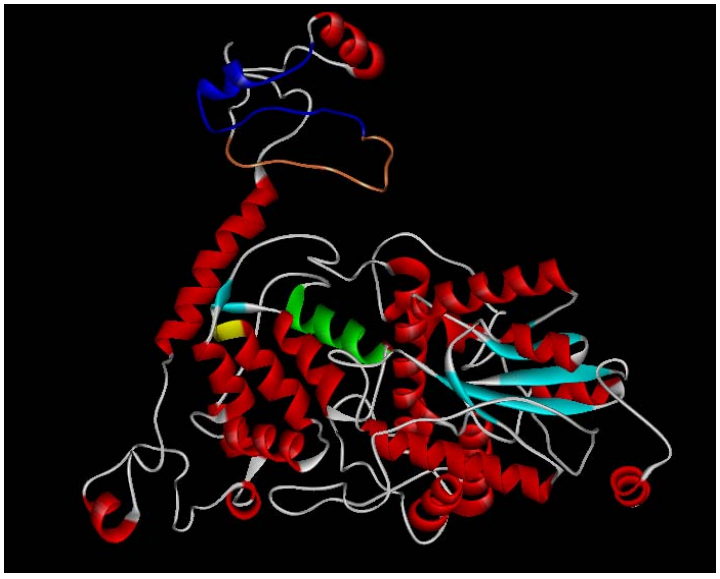


Figure 6.1: The possible RGD, Lipoprotein Attachment, Stathmin and Autophosphorylation regions of hCry2, represented with yellow, gren, blue and orange colors, respectively.

Table 6.2: The active sites and post-translational sites of hCry2 found by ScanProsite and MotifScan tools. The numbers at the left column indicate the region of protein, at which the sites are found (the amino acid numbers at the primary sequence of hCry2).

Location	**ScanProsite**
no locations given	N-myristoylation site
no locations given	Casein kinase 1E phosphorylation site
no locations given	Protein kinase C phosphorylation site
no locations given	Tyrosine sulfation site
no locations given	cAMP- and cGMP-dependent protein kinase phosphorylation site
no locations given	ATP/GTP-binding site motif A (P-loop)
no locations given	Cell attachment sequence
no locations given	Amidation site
Location	**MotifScan**
437-440	AMIDATION
347-354	ATP_GTP_A
296-299	CAMP_PHOSPHO_SITE
at 6 locations	CK2_PHOSPHO_SITE
at 6 locations	MYRISTYL
at 7 locations	PKC_PHOSPHO_SITE
411-421	Lipoprotein Attachment
386-388	RGD
333-405	CALPAIN_CAT
22-189	DNA_photolyase
230-507	FAD_binding_7

The novel sites of human Cry2 ,such as RGD, stathmin, lipoprotein attachment site, autophosphorylation site, were not found on *E. coli* DNA photolyase and *A. thaliana* Cry, as expected. Mouse Cry2 protein has exactly the same sites with the human Cry2 protein. Please refer to Supplementary # 8 for more detailed results.

The molecular function, cellular component and biological processes of the hCry2 were determined by *JaFa* [46] tool. The results are given in Table 6.3.

The *JaFa* tool predicted that;

- hCry2 may have DNA photolyase activity → it has similar sequence to DNA photolyase proteins, but no functional similarity. It has been experimentally shown that Cry does not have photolyase activity [1]. Consequently, this result is false-positive.
- hCry2 can bind to DNA → it acts as transcription regulator, so it really binds to DNA
- Protein may have protein-binding ability → hCry2 is known as a sticky protein that is able to bind many proteins
- It may have carbon-carbon lyase activity.
- hCry2 may take place in a G-protein coupled signal transduction, as we propose.

Cryptochromes of mammalians were shown to shuttle between nucleus and cytoplasm [1]. The cellular localization results of JaFa implied that the protein can shuffle between nucleus and mitochondrion. It is known the hCry2 can localize in nucleus, but localization in the mitochondria with higher score (1.6) can be seen, indicating that this protein may take place in some unknown biological pathways.

The predicted biological processes results of JaFa are about light perception and rhymtmic behaviour. The tool predicted that protein has DNA repair activity, with the highest score (3.2); but hCry2 does not have repair activity. This false-positive result is most possibly due to the high sequence similarity between *E. coli* DNA photolyase and cryptochromes.

Table 6.3: The *JaFa* results for hCry2. This tool makes a search in GO database and determines the molecular function, cellular component, biological process of the protein according to its primary sequence. The *Score* at the very right column implies the confidence level of the predcited result and the maximum score is 4. On the very left column of the table, the GO accession number of the given processes are listed.

GO ACC	GO Name	GO Root	Score
GO:0003913	*DNA photolyase activity*	*molecular_function*	*3.2*
GO:0003677	*DNA binding*	*molecular_function*	*1.8*
GO:0008020	*G-protein coupled photoreceptor activity*	*molecular_function*	*1.6*
GO:0016834	*other carbon-carbon lyase activity*	*molecular_function*	*1.2*
GO:0005515	*protein binding*	*molecular_function*	*0.4*
GO:0005739	**mitochondrion**	**cellular_component**	**1.6**
GO:0005634	**nucleus**	**cellular_component**	**0.6**
GO:0006281	*DNA repair*	*biological_process*	*3.2*
GO:0007601	*visual perception*	*biological_process*	*1.6*
GO:0007623	*circadian rhythm*	*biological_process*	*1.2*
GO:0009583	*detection of light stimulus*	*biological_process*	*0.8*
GO:0007622	*rhythmic behavior*	*biological_process*	*0.4*

The cellular localization of the protein was also examined by *PSORT II* [47]. The result is given in Table 6.4. According to *PSORTII* [47] results mitochondrial localization of hCry2 has higher probability than the nucleus localization, as the *JaFa* tool implied. The nucleus localization signal site of hCry2 protein is located at acids 559-565 region.

The *PSORT II* [47] results imply that hCry2 may localize outside the cell (8.7 %); this result supports that protein may have cellular attachment (*RGD*) region as predicted by *MotifScan* and *ScanProsite* tools. There is no experimental evidence or hypothesis about extracellular localization or function of cryptochromes; consequently this prediction of *PSORT II* [47] can be false-positive, or the cryptochromes may have unidentified roles.

It is also predicted that protein may have localization in cytoskeleton; supporting that protein may have *Calpai*n and *Stathmin* motifs and playing role in cytoskeletal functions.

Table 6.4: The *PSORT II* results for hCry2 localization. The percentages at the left column indicates the confidance level of localization at the sites given at the right column.

Percentage	Localization
47.8	Cytoplasmic
21.7	Mitochondrial
8.7	Extracellular
4.3	Cytoskeletal
4.3	Nuclear
4.3	Peroxisomal
4.3	Endoplasmic Reticulum

Secondary Structure Analysis:

The secondary structure of hCry2 was achieved by three different prediction tools; *Rosetta-HMMER* [29], *Rosetta* of Baker's lab [30] and *PredictProtein* [48]. The results of the predictions are given in Supplementary # 10.

The N-terminal (the first 480 amino acids) secondary structure predictions of hCry2 coming from all tools were almost the same. The alpha-helices and beta-strands were predicted at the same locations, diverging only 1-2 amino acids.

The C-terminal (the last 100 amino acids) predictions of the tools were different from each other. The Rosetta-HMMER [29] tool predicts one alpha-helix at amino acid *492-510*; and one beta-strand at amino acid *574-581*. According to Rosetta server human Cry2 C-terminal has two alpha-helices at amino acid *491-510* and *574-578*. The PredictProtein tool predicted alpha-helix at amino acid *493-512, 559-561, 572-575,* and beta-strand at amino acid *517-520*.

Rosetta tool predicts a large portion of the C-terminal (from amino acid *511 to 573*; and from amino acid *579 to 593*) as "disordered", meaning that this part of the protein is unstructured. The other secondary structure tools also estimates large unstructured regions

at approximately the same locations. These data in agreement with recent paper stating that the NMR analysis implies the structure of last 100 amino acids of Cry2 likely to have a alpha-helice at the beginning and then the structure becomes unstructured [72].

The secondary structure indicates that the N-terminal of hCry2 is similar to DNA photolyase and known Cry2 structures, but has an extra alpha-helix at the very beginning. The C-terminal of the protein has a large alpha-helix at domain *491-510*, and a beta-strand or an alpha-helix at the amino acid *572-581* region.

Tertiary Structure Analysis:

The *Robetta* [36] tool was used for *ab initio* modeling; while *EsyPred* [34] and *Consensus* [35] tools were employed for comparative modeling. The primary sequence of hCry2 was given as query to all modeling tools. The results from *Consensus* and *EsyPred* were obtained in a short time, while the 10 results from Robetta obtained after 5 months from the submission date.

The 3D model made by *Consensus* (Figure 6.2) is from the amino acid # 21 (Serine) to amino acid # 500 (Isoleucine). The *EsyPred model* (Figure 6.3) is from amino acid # 23 (Serine) to amino acid # 508 (Glutamine). Both of these models lack the C-terminal part of Cry2. This is an expected result, since there is no known homolog structure to the C-terminal of hCry2.

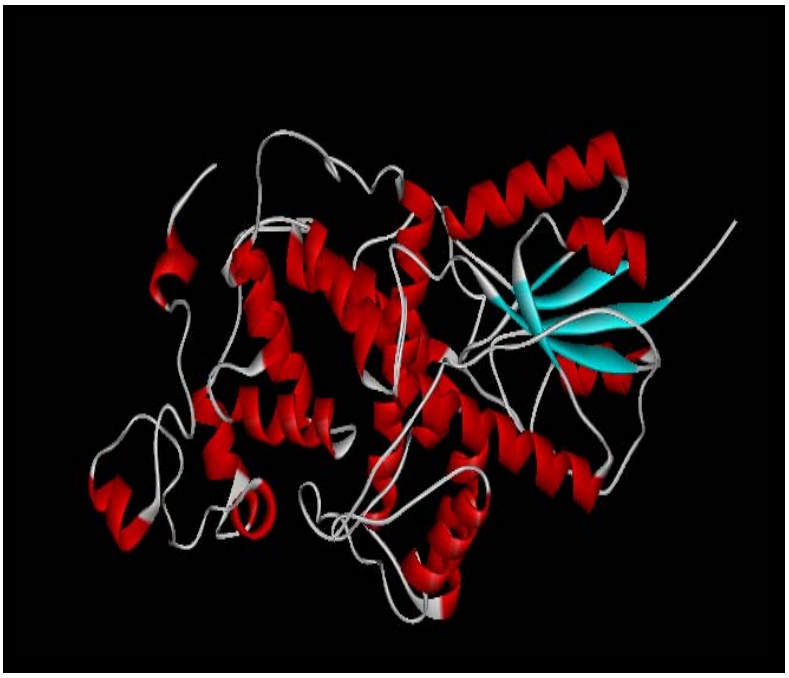


Figure 6.2: The model of hCRY2 done by Consensus Modeler. (by WebLab ViewerPro Version 4.0)

Figure 6.3: The model of hCRY2 done by EsyPred Modeler. (by WebLab ViewerPro Version 4.0)

The 10 models done by *Robetta* server is from amino acid # 1 (methionine) to amino acid # 593 (alanine). The N-terminal of the models was similar to DNA photolyase, and structurally known Cry2 proteins. This was an expected result, since *Robetta* server firstly makes homology search. The difference of hCry2 N-terminal from other organisms' Cry2 proteins is the alpha-helix at the very beginning of the protein.

A part of C-terminal (starting from residue # 511) was made by *ab initio* modeling. The C-terminal for 10 models was different from each other; but every model has one or two alpha-helices. The remaining parts of the C-terminal were all coil. The unstructured C-tail estimation of secondary structure prediction tools and Partch et al [72] experimental results is supported by those 3D structure predictions.

Refinement of tertiary Structure: The most reliable model among 10 models of *Robetta* was determined by the "Protein Model Check" and "Ramachandran Analysis" options of *WhatIf* [37]. The packing quality, Ramachandran plot analysis, backbone conformation, bond angles, bond lengths, side-chain planarity, omega angles of hCry2 models were calculated by the tool. The results of these runs were compared with cut-off values. The experimental structures for *A. thaliana* Cry and *H. influenza* universal stress proteins taken from *PDB* were also run by *WhatIf* tool, as positive-controls (Please refer to "Supplementary # 9"). The results implied that the **model # 3** is the most stable among all 10 models. Consequently the 3rd model was chosen for docking studies and for further analysis.

Model # 3 was compared with other *Robetta* models and the secondary structure predictions (Please refer to "Supplementary # 10"). The comparisons indicated that there are secondary structure motifs of model # 3, which are under or over estimated. Some corrections have been made on model # 3; and its .pdb file and the apropriate coordinates were taken from the other 9 models of *Robetta*. The altered residues, the type of secondary

structures of those residues before and after the alterations; and the model number from which the new coordinates were taken are summarized at Table 6.5.

Table 6.5: The first column respresents the number of residues that alteration was done; the second column indicates the secondary structure of that region before the alteration (the wrongly estimated structure) and the third column shows the secondary structure after the alteration; the last column is for the Robetta model # from which the new coordinates are taken.

Residue Number	The structure before alteration	The structure after alteration	Alteration made by taking from the model number:
from 12 to 15	alpha-helix	coil	4
from 254 to 258	alpha-helix	coil	4
from 301 to 302	alpha-helix	coil	4

The comparison of the secondary structures of models starting from residue # 511 (the part of protein that was done by *ab initio* technique) with the results of secondary structure prediction tools imply that the most possible structure is from **model # 6** (please refer to Supplementary # 10). Consequently, the modified model # 3 (residues from 1 to 510) were combined with the C-tail of model # 6 (residues from 511 to 593).

The final model was refined by “remove bumps” and “add side-chains” options of *WhatIf* tool (Figure 6.4).

The predicted model of hCry2 has two domains, connected with a loop structure. The first domain is the *alpha/beta domain*, which is constructed by 6 alpha-helices and 5 beta-sheets. The domain consists of 145 residues. A long *loop (coil) structure* (connector region) follows the alpha/beta domain. This part is from residue # 146 to residue # 233. An alpha-helix sits in this loop part between residues 170-180. There is a 3/10 alpha-helix sitting in this region between the residues 153 (ASP) – 156 (ARG). The connector region joins the alpha/beta domain to the second domain of hCRY2; the *alpha domain*. In the alpha domain

there are 15 alpha-helices and 2 short beta-sheets (between residues 339-340 and 390-391). This alpha domain of hCry2 also has five 3/10 alpha-helices.

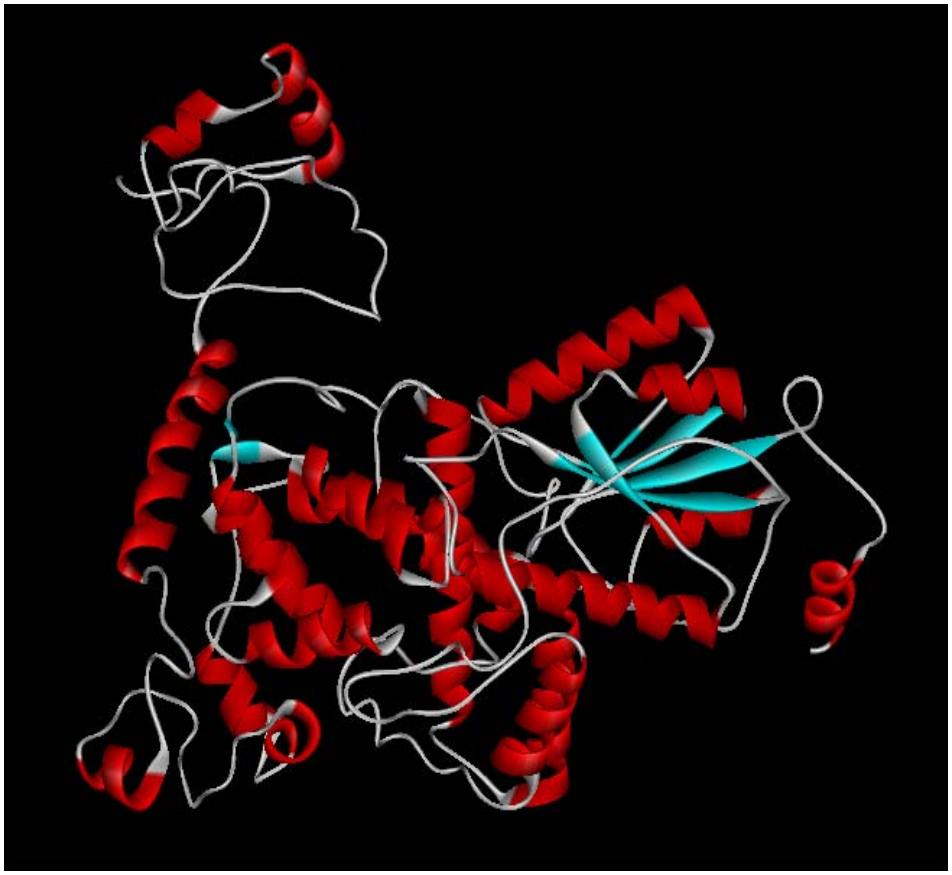


Figure 6.4: The predicted tertiary model of hCRY2 by Robetta server and refined. (by WebLab ViewerPro Version 4.0)

The *A. thaliana* Cry should have been used as the template for homology modeling of hCry2, but the structure of the protein should not be the same of *A.thaliana* Cry; since the sequence of hCry2 is different than *A. thaliana* Cry sequence. The reliability of models should have been checked, in order not to have a copy-paste of a known protein. Consequently, the hCry2 models from Robetta [36], EsyPred [34] and Consensus [35] are

superimposed with *A. thaliana* Cry via *FatCat* [33] tool (Table 6.6). As can be seen from the results; the hCry2 model is not the exact match of *A. thaliana* Cry, but similar to it, as expected. Consensus predicts many “coil” regions, whereas EsyPred predicts “helice parts” instead of it. EsyPred model is more similar to *A. thaliana* Cry than the Consensus model. The most distinct model from *A. thaliana* Cry is the Robetta model.

Table 6.6: The superimposition results of hCry2 models by *A. Thaliana* Cry2 protein. *A. Thaliana* Cry2 is 485 amino acids long protein. The match length of each superimposition and the rmsd value are given. The Robetta model is the most distinct one among the three models.

	Robetta Model
Match Length	476
rmsd	3,15
	Consensus Model
Match Length	465
rmsd	2,66
	EsyPred Model
Match Length	465
rmsd	2,57

The *E. coli* DNA photolyase and *Synechocytis* Cry 3D structures should be similar to structure of hCry2, and they could have been used as template of modeling of the protein. The reliability of hCry2 model was checked by superimposition of the protein model with *E. coli* DNA photolyase and *Synechocytis* Cry. The results are given in Supplementary # 14.

The 3D similarity search of the hCry2 model: The tertiary model of hCry2 (Figure 6.4) was used as input to find other proteins that are structurally similar to itself. The

superimposition of protein to the database was achieved by *EMBL-Dali* [42]. The results of this tool deciphered probable novel functions of the human Cry2.

The *EMBL-Dali* [42] results imply that the N-terminal of protein may have the chaperone activity (Table 6.7). The 3D structure of chaperone protein; the UspA (*PDB* id: 1jmv-A) is obtained from Protein Data Bank. UspA is made up of 4 identical domains.

Table 6.7: The first 6 results of *EMBL-Dali* search of Robetta model of hCry2. Global alignment of proteins are employed by the tool.

Z-Score	RMSD	% Identitiy	Protein Name	pdb Code
36,00	2,2	27	DNA REPAIR photolyase	1qnf
9,2	**3,7**	**10**	**CHAPERONE universal stress protein a (uspa)**	**1jmv-A**
9,10	3,6	11	atp-binding domainof protein, hypothetical	1jmh-A
8,00	3,5	10	hypothetical pro	1q77-A
7,80	4	15	putative n-type	1ru8-A
7,80	4,3	12	electron transfer flavoprotein	1efv-A

The superimposition of hCry2 N-terminal and the domains of UspA was achieved by *FatCat* [33] and *CE* [32] tools. The results of superimposition imply that there is a little difference between hCry2 and all 4 domains. (Please refer to “Supplementary # 11”). The least rmsd (2.76) difference is observed between hCry2 N-terminal and UspA domain # 3. The gap number, that is the resdiues that do not match, is 25. The difference in the length of the two proteins is 17 amino acids; consequently the real gap is only **8 residues**. All of the alpha-helix and beta-strand parts of two proteins match, except the very first alpha-

helix of hCry2; and the alpha-helix of UspA at residues 53-58. These results indicate that hCry2 N-terminal may have a chaperone activity (Table 6.8 and Figure 6.5).

Table 6.8: The superimposition result of UspA domain # 3 and N-terminal of hCRY2 Robetta model.

UspA Length	133 amino acids
hCRY2 N-terminal Length	150 amino acids
Matching Length	109 amino acids
rmsd	2,76
Gaps	25,00

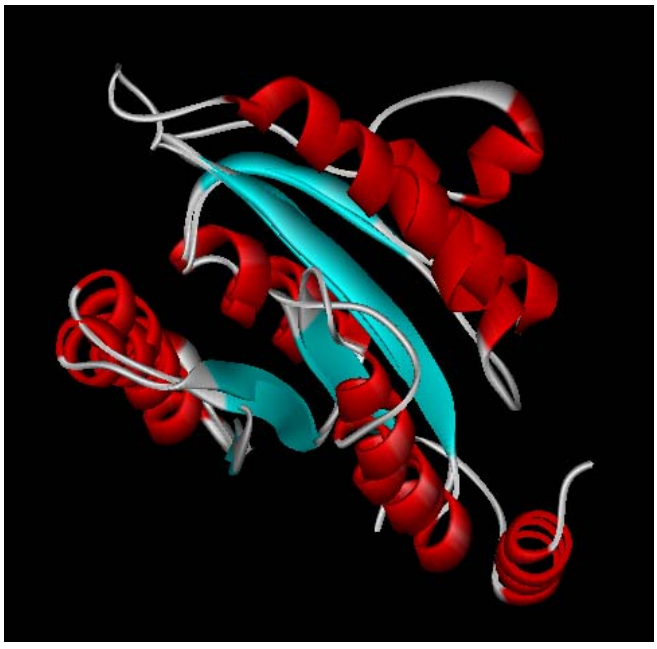


Figure 6.5: The superimposition of UspA and hCRY2 N-Terminal. (Please refer to “UspA_CRY2.pdb” file to examine the superimposed proteins)

6.1.2 Melanopsin Protein

Hamster and human melanopsins primary amino acid sequence similarity was determined via *BLAST* [28] algorithm, and they were found to share a homology (Table 6.9). Please also refer to Supplementary # 12.

Table 6.9: H. sapiens versus P. Sungorus (hamster) melanopsin *BLAST* result. The gap, that is the difference between sequences is 3 %; the Expect (E-value) is very small meaning that the sequences are very similar; the identical residues are 75 % of total; the resemblance is 85 %.

Expect	9,00E-177
Identities	75%
Positives	83%
Gaps	3%

This similarity allow us to predict tertiary structure of human Melanopsin. Since the Melanopsin is a transmembrane protein, the Cry2 can only reach and dock to its cytosolic parts or loops between membrane domains. Therefore the membrane, cytosolic and extracellular parts of this protein were determined by the *SPLIT* [40]. The results are given in Table 6.10.

Table 6.10: The transmembrane parts of Melanopsin predicted by *SPLIT*. The first column represents the number of the transmembrane; the second and third columns indicate the amino acid number that the alpha-helice starts and ends.

# of transmembrane alpha-helice	starts at amino acid #	ends at amino acid #
1	70	93
2	104	128
3	143	167
4	187	207
5	233	260
6	292	323
7	357	373

According to the *SPLIT* [40] results Melanopsin has three cytosolic loops between residues; 94-103, 168-186, 261-291 and the C-tail strating at residue number 373, and ending at amino acid number 469, the final residue of the protein. The sequence similarity results between those cytosolic regions of hamster and human Melanopsin were compared (Please refer to Supplementary # 12). The different residues were determined; those different residues of hamster Melanopsin were altered by its corresponding residue in human Melanopsin via the “mutate a residue” option of *WhatIf* [37] tool. This option of *WhatIf* makes one mutation per one run; consequently numerous runs were done. The tertiary structure of hamster Melanopsin is given to the server (in .pdb format) as query; and the results were taken interactively. The result was again in .pdb format.

The final mutated tertiary model of hamster Melanopsin was containing the cytosolic loop residues of human Melanopsin. This model was refined by “remove bumps” and “complete a structure” options of *WhatIf* [37].

After alterations and refinements, it was possible to use the 3D model of Melanopsin for docking analysis.

6.1.3 hCry2-Melanopsin Docking

The electrostatic forces and shape complementarities between Melanopsin and hCry2 proteins had been investigated by Hex 4.5; afterwards detailed docking analysis between these proteins investigated by the 3.0.5 version of AutoDock tool, as described in Section 4.3.

Hex Docking Results:

The probable interaction between human Melanopsin-Cryptochrome; and also the possible interactions Cryptochrome makes with itself were investigated via “Shape+Electrostatic Search Docking” option of *Hex 4.5* tool, as described in Section 4.3.1.

As implied in Section 6.1.1; it was found that C-terminal of protein has an unstructured shape. The C-tail is a very large region to be unstructured (approximately 100 amino acids), consequently we propose that it should have ability to fluctuate between structured and unstructured states in order to be able to function. The structure of alpha/beta domain of hCry2 was found to be similar to Universal stress protein a (UspA) with chaperone activity, as described in Section 6.1.1. Consequently, it is possible that N-terminal of hCry2 has chaperone activity. This N-terminal may dock to the unstructured C-tail; making it able to take a structure by its probable chaperone effect. To analyze this probable interaction, the N- and C-terminals of hCry2 were investigated by Hex 4.5.

To determine whether UspA and hCry2 have only structural similarity or also functional similarity; UspA was docked to the cytosolic part of the Melanopsin. As a control, another structurally UspA similar protein (Phosphoantethein adenylyl transferase, *PDB* id: 1B6T) with no chaperone activity was also docked to cytosolic part of Melanopsin. If hCry2 is structurally and also functionally similar to UspA, then UspA should dock to Melanopsin with a similar energy (compared to hCry2) and 1B6T should dock with more positive energy.

As implied in Chapter 4, the following dock searches were performed;

1. hCry2 → receptor; Casein kinase 1E → ligand (known to dock, used as control).

Melanopsin and Cryptochrome interaction?

2. hCry2 → receptor; cytosolic part (all three loops and C-tail) of Melanopsin → ligand
3. hCry2 → receptor; first cytosolic loop of Melanopsin → ligand.
4. hCry2 → receptor; second cytosolic loop of Melanopsin → ligand.
5. hCry2 → receptor; third cytosolic loop of Melanopsin → ligand.
6. hCry2 → receptor; C-tail of Melanopsin → ligand.

Controls

7. *A. thaliana* Cry → receptor; cytosolic part of Melanopsin → ligand (negative control)
8. *E. coli* DNA Photolyase → receptor; cytosolic part of Melanopsin → ligand (negative control).

Homodimerization occurrence?

9. hCry2 N-terminal → receptor and ligand
10. *A. thaliana* Cry N-terminal → receptor and ligand (known to dock, used as control for hCry2 homodimerization).

Chaperone activity?

11. hCry2-UspA similar part → receptor; hCry2 C-tail → ligand.

Controls:

12. UspA → receptor; cytosolic part of Melanopsin → ligand.
13. 1B6T protein → receptor; cytosolic part of Melanopsin → ligand.
14. hCry2-UspA similar part → receptor; cytosolic part of Melanopsin → ligand.

Table 6.11 summarizes the dock energies taken from Hex 4.5 simulations. Since *A. thaliana* Cry is known to make a homodimer; its resulting dock energy (-0.57) is chosen as cut-off value for Hex 4.5 tool. The possible interacting protein pairs' dock energies are colored with blue. The color of the pairs that are not possible to interact is yellow.

The Hex 4.5 tool was one more time shown to be reliable by the control run; Casein kinase-Cryptochrome were shown to dock with interaction energy of -1.18746.

Melanopsin and Cryptochrome interaction: According to Hex 4.5 results (Table 6.11); the first loop and the third loop of Melanopsin interact with hCry2. The interaction between the first loop-Melanopsin (-0.741809) is stronger than the interaction between the third

loop-Melanopsin (-0.597738). Also the C-tail of Melanopsin binds to hCry2 with very strong energy (-1.385114).

The interaction between hCry2-Melanopsin is also verified by the negative controls; as expected the *A. thaliana* Cry-Melanopsin interaction (-0.564674) and *E. coli* Photolyase-Melanopsin interaction (-0.380849) are not possible when their dock energies are compared to the dock energy of hCry2-Melanopsin, -0.752077 and cut-off value -0.57.

The Hex 4.5 results imply that the interaction between Cryptochrome-Melanopsin is possible; and the interaction is via the 1st and 3rd loops and C-tail of Melanopsin. The interaction between cytosolic loops of Melanopsin and hCry2 were studied further with AutoDock tool.

Homodimerization occurrence: It is known that *A. thaliana* Cry makes homodimer [45], via its N-terminal to start signal transduction. The Hex 4.5 result implies that the interaction between N-terminal of hCry2 with itself is very possible (Table 6.11); since the control (*A. thaliana* homodimer) interaction energy was found to be -0.57315, whereas the interaction energy of hCry2 N-terminal to itself was calculated as -0.79203. The possible interaction is shown at Figure 6.6. As can be seen from Figure 6.6, the protein makes homodimer with symmetry.

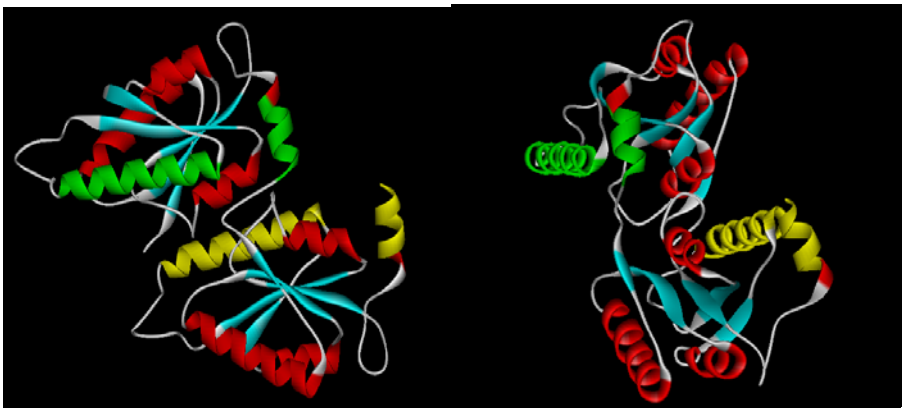


Figure 6.6: The possible homodimer formation of hCry2 N-terminals is given in 2 different perspectives. Some regions of the one domain were colored with green color, while the same regions on the second domain were colored with yellow in order to clarify the symmetry of interaction.

Table 6.11: Results of Hex 4.5 dock simulations.
The ligand --- receptor names are written at the 1st column; the total energy (electrostatics and Van der Waals), which is calculated by "Energy.py" script is written at the 2nd column.

Control:	
	Energy (kJ/mol)
Casein Kinase 1E --- Cry2	**-1.187**
Melanopsin and Cryptochrome interaction?	
Melanopsin --- Cry2	**-0.752**
Melanopsin 1st loop --- Cry2	**-0.742**
Melanopsin 2nd loop --- Cry2	-0.303
Melanopsin 3rd loop --- Cry2	**-0.598**
Melanopsin C-tail --- Cry2	**-1.385**
Controls for this interaction:	
Melanopsin --- A. thaliana Cry	-0.565
Melanopsin --- E. Coli DNA Photolyase	-0.381
Homodimerization occurrence?	
Cry2 N-terminal to itself	**-0.792**
A. thaliana Cry N-terminal to itself	**-0.573**
Chaperone activity?	
Cry2-UspA Similar part --- Cry2 C-tail of model #3	**-0.576**
Cry2-UspA Similar part --- Cry2 C-tail of model #6	-0.406
Controls for this interaction:	
Melanopsin --- UspA	**-1.394**
Melanopsin --- 1B6T	-0.311
Melanopsin --- Cry2-UspA Similar part	**-0.589**

Chaperone activity: The Hex 4.5 results (Table 6.11) imply that the interaction between hCry2-UspA similar domain and Melanopsin (-0.58895) is weaker than the interaction between UspA and Melanopsin (-1.394069); but it is stronger than the interaction between 1B6T (structurally Usp similar protein) and the Melanopsin (-0.311212). Those results imply that UspA is not only structurally but also functionally similar to N-terminal of hCry2. The structural and functional similarity between those proteins implies that hCry2 can have chaperone activity.

It is known that hCry2 C-tail is unstructured, so the C-tail of this protein can interact with its UspA similar domain and it can take a structure, as stated before.

The most stable model from Robetta Server was found to be the model # 3 by *WhatIf* server. The model # 3 was refined as described in Section 6.1.1 (the C-tail of model # 6 was used for refinement). When the C-tails of model # 3 and the refined model are examined, it is observed that the C-tail of model # 3 has one 1-turn and one 2-turn alpha-helix at residues 516-522 and 524-528; and refined final model has 2 alpha-helices at residues 561-565 and 572-582.

What is suprising is that one model has coils that are structurally very similar to alpha-helices at almost the exact regions that the other model has alpha-helices (Table 6.12 and Figure 6.7).

Table 6.12: The coils, which are similar to alpha-helices, and alpha-helix regions of the most stable *Robetta* model and final model are given. It is obvious that the alpha-helix regions in one model are a coil region in the other model.

	Robetta Model # 3	**Final Model**
Alpha-helix	*516-522*	**561-565**
Alpha-helix	*524-528*	**572-582**
Coil similar to Alpha-Helix	**557-565**	*518-522*
Coil similar to Alpha-Helix	**572-583**	*522-528*

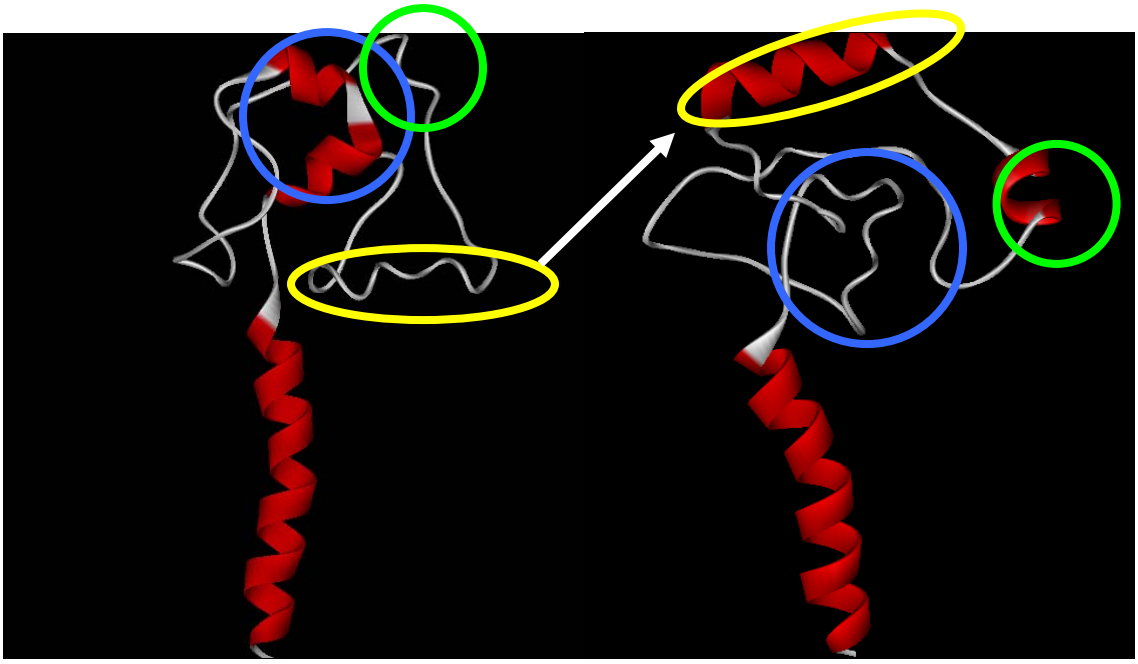


Figure 6.7: The C-tails of Robetta model # 3 (at the left) and the final model (at the right) of hCry2 are shown.
The alpha- helices at the left disappear at the right; and the probable alpha-helices formed at the right, and vice-versa. The probable alpha-helix and corresponding alpha-helix are indicated with the same colored circles; and one of them is shown with an arrow as an example.

Table 6.12 and Figure 6.7 imply that there can be a switch between the C-tail structures of model # 3 and refined final model. According to WhatIf tool model # 3 of Robetta is energetically more stable than model # 6; but the secondary structure predictions imply that the C-tail should be similar to the C-tail of the model # 6. Those secondary structure predictions and energy terms also support the idea of switch between 2 models.

The final model has 2 alpha-helices one of them between 561-565 and the other at 572-582. The 559-565 residues of hCry2 was found to contain “Nuclear Localization Signal” via the *PSORT II* tool. The 553 and 557 residues of mouse Cry2 (mCry2) are found to contain phosphorylation site by Harada et al [50]; and the phosphorylated protein was

found to be localized in the nucleus; after the phosphorylation degradation takes place. Since the mouse and human Cry2 proteins have a protein sequence only varying by 1 residue; the same phosphorylation/degradation mechanism can occur in humans, too.

According to those information, the C-tail of model # 3 can make hCry2 localize in cytoplasm; afterwards the C-tail of model # 6 can drive the protein to nucleus. For the switch between two structures of C-tail to occur, the possible chaperone domain of hCry2 may play role.

In order to be able to detect whether interaction between C-tail and UspA domain of hCry2 occurs or not Hex 4.5 tool was used (Table 6.12). When the C-tail of hCry2 of model # 3 was used as ligand and the UspA domain of hCry2 was used as receptor, the result (-0.57599) implied that this interaction is possible. Whereas, when the C-tail from model #6 was used as a ligand, the interaction to UspA domain was observed to be impossible (-0.40582) according to the cut-off value.

AutoDock Results:

The *Hex 4.5* tool was used as a preliminary tool for hCry2-Melanopsin interaction. The results of *Hex* imply that the interaction between hCry2-Melanopsin is possible. Consequently, detailed analysis of this probable interaction was achieved via *AutoDock* tool version 3.0.5.

As stated in Section 4.3.2, *AutoDock* tool does not have the ability to accept proteins longer than approximately 10 amino acids as ligand due to its a rotatable bond limit. Consequently, the full Melanopsin protein could not be used as ligand. Since the cytosolic loops of the protein are short, the loops were used as ligand firstly. The 1st loop was from residue 97 to 104; the 2nd loop was partitioned into 2 subgroups from 167-176 and 177-183; the 3rd loop was from 267 to 276. Secondly, the C-tail of Melanopsin was partitioned into small portions and used as ligand. The portions were as follows; 373-383, 384-394, 395-405, 406-416, 417-427, 428-438, 439-449, 450-460, 461-468.

hCry2 was used as receptor. Since it is a bulky protein, it could not be able to fit in a grid of dimensions 126x126x126 Angstrom^3. Consequently, 5 different grids were formed, as implied in Chapter 4. The center of each grid in x-y-z format is as follows;

Grid #1 → 36, 3, 5.4
Grid # 2 → 18, -8, 5
Grid # 3 → 13, 33, 5.4
Grid # 4 → 15, -10, 5
Grid # 5 → 40, 3, -10.

The ligands and receptor were prepared for docking according to the method given in Section 4.3.2. The grid and docking preparation is also described in Section 4.3.2.

The results of *docking energy* of hCry2-Melanopsin are given at Table 6.13. The energy is in terms of *kcal/mol*.

Table 6.13: The docking energies (kcal/mol) of hCry2 and Melanopsin cytosolic loops & C-tail are given. The 1st column indicates the portions of Melanopsin that were used as ligand. The proceeding columns indicate the docking energies of the **hCry2 part in the indicated grid** and **Melanopsin portion**. The most possible interaction for each Melanopsin portion is indicated with ***italic***.

	Grid # 1	**Grid # 2**	**Grid # 3**	**Grid # 4**	**Grid # 5**
1st Loop	-11.25	-12.34	-11.39	-11.01	***-12.68***
2nd Loop	-11.34	-11.7	-10.87	***-13.17***	-10.7
3rd Loop	-10.96	-12.1	-10.54	***-13.26***	-9.96
373-383	-11.12	-12.26	-10.92	***-13.31***	-9.2
384-394	-9.71	***-11.02***	-9.24	-10.33	-8.36
395-405	-13.45	-13.39	-11.9	***-13.59***	-10.95
406-416	-11.55	***-14.1***	-9.86	-13.72	-11.89
417-427	-11.55	***-13.49***	-10.23	-10.56	-10.86
428-438	-11.56	***-14.85***	-13.12	-14.76	-12.93
439-449	***-13.77***	-12.68	-12.17	-12.75	-13.16
450-460	-10.73	***-17.1***	-12.16	-15.93	-11.53
461-468	-12.15	***-13.52***	-10.89	-11.33	-9.75

Only the Grid # 3 contains the C-tail of hCry2, other grids contain different parts of the N-terminal of the protein. As can be seen from the Table 6.13, none of the portions of Melanopsin did bind to the C-terminal (Grid # 3) of hCry2. Consequently, the interaction between cytosolic loops of Melanopsin and hCry2 is via the N-terminal of hCry2.

The C-tail of Melanopsin was found to dock to the N-terminal of hCry2 with low energy values. Escepially, the region between 450-460th amino acid residues of Melanopsin binds to hCry2 with the lowest energy value (-17.1 kcal/mol).

According to the results of AutoDock, a large portion of Melanopsin cytosolic loops and tail interacts with N-terminal of hCry2.

The possible interactions, via the Melanopsin cytosolic loops, are shown at Figure 6.8. The 2nd and 3rd loops seem to dock to the same region of the hCry2 protein. There are two possibilities; the region that they are binding can have the ability to open and close (this region is a kind of cavity, so it is possible for it to change conformation) or 2nd loop may not bind to hCry2 as Hex 4.5 results also imply. But the hCry2 region that 1st loop binds to is different than the region that the other 2 loops bind to. This region is exactly the UspA domain. The 1st loop may bind to this region and it may also take a secondary structure due to the probable chaperone activity of hCry2.

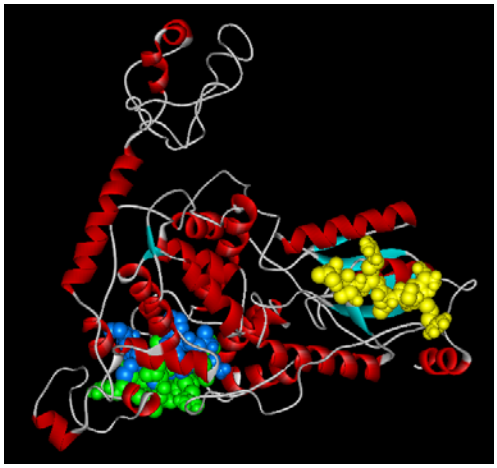


Figure 6.8: The interaction between cytosolic loops of Melanopsin (yellow → 1st loop; green → 2nd loop; blue → 3rd loop) and hCry2.

AutoDock was also used to find the interaction between hCry2 and its cofactors; FAD and MTFH. The interaction energies are given at Table 6.14.

Table 6.14: The interaction energies (kcal/mol) of hCRY2 and its cofactors are given.

	Grid # 1	**Grid # 2**	**Grid # 3**	**Grid # 4**	**Grid # 5**
FAD	-13.93	***-16.08***	-13.51	-13.9	-12.54
MTFH	-12.2	-12.64	-11.35	***-13.25***	-10.73

According to Table 6.15; FAD and MTFH bind to the N-terminal of the protein. The interaction between FAD-hCry2 is stronger than the interaction between MTFH-hCry2. This was an expected result, since MTFH acts as antenna and FAD is known to be embedded in *E. coli* Photolyase and *A. thaliana* Cry proteins. The interactions are shown at Figure 6.9.

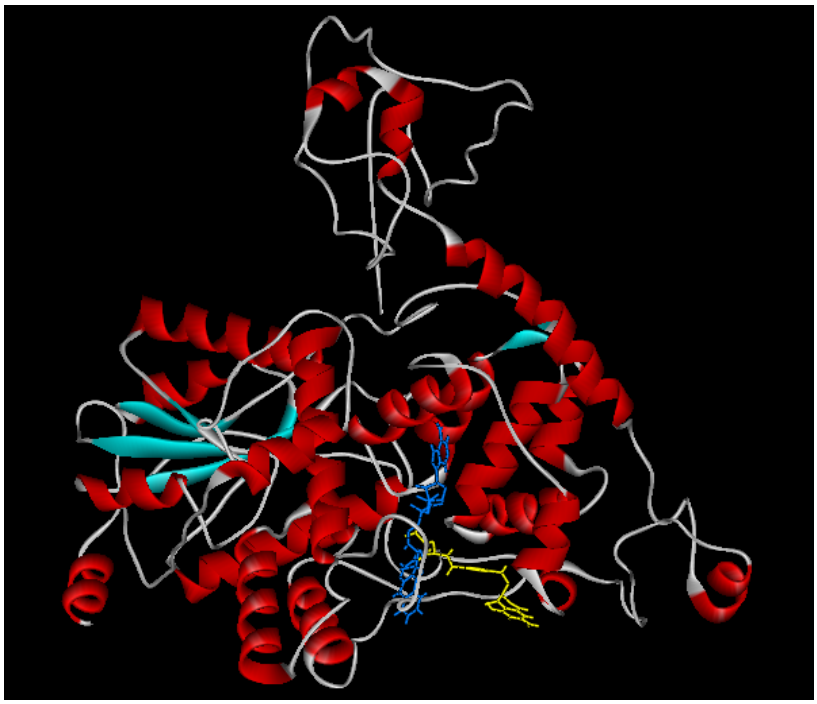


Figure 6.9: The interaction between FAD (blue), MTFH (yellow) and hCry2.

The Residues at the Interaction Surface:

The residues at the interaction surface of *AutoDock* results were analyzed by *NACCESS* [66]. The absolute accessible surface areas of hCry2-Melanopsin pairs and the absolute surface area of hCry2 alone were calculated. The accessible surface areas of hCry2 residues (hCry2 alone, hCry2 interacting with 1st loop, hCry2 interacting with 2nd loop, hCry2 interacting with 3rd loop, hCry2 interacting with all C-tail portions) were taken from corresponding *.rsa* file.

The surface area of hCry2 alone was compared with the surface area of hCry2 in an interaction with Melanopsin. If one of the residues' absolute accessibility value in protein complex is decreased by >=1 Angstrom^2, then that residue was determined to be at the interface of the complex (Please refer to Supplementary # 13).

The results are summarized at Table 6.15. The residues of hCry2 at the interaction surface are given. Some of the hCry2 residues are at the interaction surface of almost all of the hCry2-Melanopsin interactions; while some of them are found only at one interaction pair.

The residues 228-273, 310, 313, 353-380, 428-455 of hCry2 interact with almost all of the Melanopsin portions. The Cryptochromes are known as sticky proteins; consequently those resdiues may be the ones that lead to stickiness of the protein. Consequently, the interactions consisting only those regions may be false-positives due to the stickiness of hCry2. On the other hand they may be real interactions. In oder to determine whether those interactions are real, experimental clarifications have to be carried out (like yeast two-hybrid) with help site-directed mutagenesis.

The interaction of hCry2 with the 1st loop, residues 395-405 and 439-449 of Melanopsin is via different residues of hCry2.

Table 6.15: The interface residues of hCry2 protein are given, i.e. the first column indicates the residues of hCry2 interacting with 1st loop of Melanopsin. The residues that are that are found in many interaction surfacesare written in **bold**. The residues that are found only at one type of interaction are written in *italic*.

1st loop	2nd loop	3rd loop	373	384	395	406	417	428	439	450	461
59-61	**228-229**	**228-229**	**228-229**	**226-229**	*171-172*	**226-231**	**226-231**	**224-229**	*65*	**226-228**	**224-228**
97-100	**231**	**231**	**231**	**231**	*175-176*	**241**	**241**	**231**	*70-71*	**245**	**245**
103-104	**245**	**245**	**245**	**245**	*179-180*	**245**	**244-245**	**241**	*73-74*	**251**	**251**
107-108	**251**	**251**	**251**	**254-255**	**256**	**249**	**249**	**245**	*211-213*	**254-258**	**255**
110-111	**254-255**	**254-255**	**254-255**	**257-258**	**262-266**	**251**	**251**	**249**	*217-219*	**263-265**	**257-258**
130	**257-258**	**257-258**	**257-258**	**260-261**	**306**	**254-255**	**255**	**251**	*221*	**267-268**	**263-264**
202-204	**263**	**260-261**	**260-261**	**263-269**	**428**	**257-258**	**257-258**	**254-255**	***223***	**272**	**267-268**
206	**268-269**	**263-264**	**263-264**	**272-273**	**430-435**	**263-264**	**268**	**257-258**	**353**	**310**	**375**
	272-273	**266-269**	**266-269**	**310**	**439**	**266-269**	**272**	**263-264**	**355**	**313**	**377**
	360	**272-273**	**272-273**	**313**	**442**	**272-273**	**360**	**267-268**	**369-370**	**373**	**443-446**
	373	**310**	**310**	**360**	*494-495*	**310**	**375**	**272**	*404-405*	**375-377**	**448-449**
	375-377	**373**	**360**	**373**	*498-499*	**313**	**377**	**360**		**380**	**451-453**
	442-446	**375-377**	**373**	**375-377**	*502*	**360**	**443-446**	**373**		**443-446**	
	448-449	**442-446**	**375-377**	**380**		**373**	**448-449**	**375-377**		**448-449**	
	452-453	**452**	**442-446**	**443-444**		**375-377**	**452-453**	**442-443**		**452-453**	
			448-449	**448-449**		**380**		**445-446**			
			452	**451-452**		**443**		**448-449**			
						452-453		**451-453**			
								455			

6.2 Results and Discussion of Experimental Methods

The computational results imply that there is a possible interaction between Melanopsin and hCry2 proteins. In order to confirm those computational results, an experimental approach was taken. The experimental procedures that are described in Chapter 5 was accomplished.

The proteins were successfully expressed in 293T cell line; cells were fixed with cover glasses and examined by fluorescence microscopy. The localizations of proteins that were tagged with fluorescent proteins were determined. The Melanopsin tagged with red fluorescent protein (RFP) was localized in cell-membrane, Figure 6.10; while the Cry2 protein tagged with green fluorescent protein (GFP) was dispersed in the cytoplasm, Figure 6.11. When the cells with both proteins; Melanopsin-RFP and Cry2-GFP were examined (Figure 6.12), it was observed that Cry2 protein was localized near transmembrane; it was no more dispersed in the cytoplasm. Those results indicate that there may be an interaction between Melanopsin-Cry2 proteins.

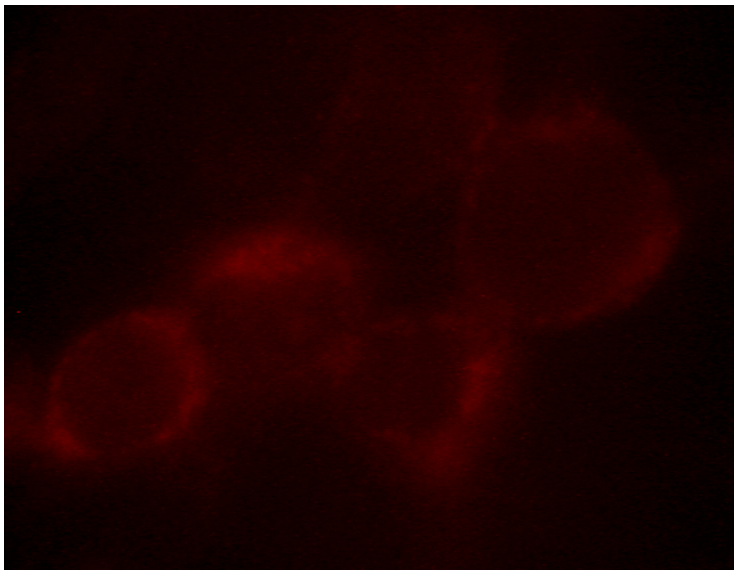


Figure 6.10: Cells transfected with Melanopsin only. Since the protein is tagged with RFP, it is visible under fluorescent microsopy. The protein is localized in the cell-membrane.

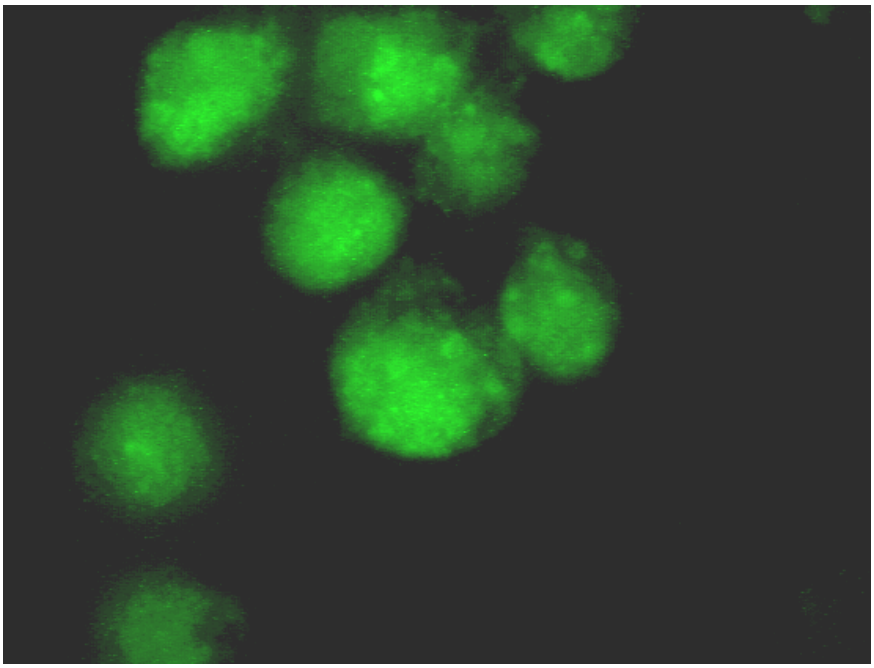


Figure 6.11: Cells transfected with Cry2 only. Since the protein is tagged with GFP, it is visible under fluorescent microsopy. The protein is localized all in the cytoplasm.

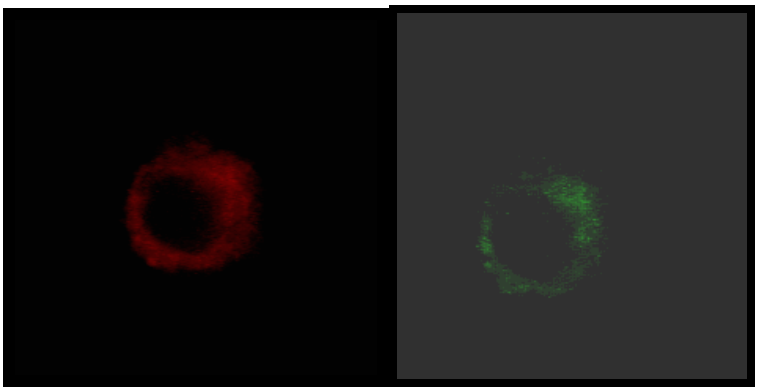


Figure 6.12: Cells transfected with Melanopsin and Cry2. On the left side Melanopsin can be seen and on the right side Cry2 can be observed. Both protein localize the cell-membrane.

Chapter 7

CONCLUSION

The computational results indicate that there is a possible interaction between Melanopsin and Cry2 proteins of mammalians. Further experimental tests, like co-immunoprecipitation and yeast two-hybrid assay, are required in addition to the tests that have been performed in this study to determine the interaction between these proteins.

The *AutoDock* [38], *Hex* [43] and *NACCESS* [66] results imply that the possible interaction between hCry2 and Melanopsin is via the N-terminal of hCry2 and 1st loop, residues 395-405 and 439-449 of Melanopsin. The NACCESS [66] tool implies that the possible interaction of proteins is via the 59-223; 404-405; 494-499; and 502 residues of Cry2 Site-directed mutagenesis experiments need to be performed for verification of these results.

Addition to Cry2-Melanopsin interaction, possible properties and computational tertiary model of hCry2 was discovered; and shed light onto possible mechanisms of Cry2 function in cell.

According to tertiary and secondary structure prediction results, hCry2 has a similar but not exact structure to *A. thaliana* Cry and to *E. coli* DNA photolyase. The protein is made up of two domains; alpha/beta and alpha connected by a large loop section. The C-terminal of the protein has a large alpha-helix part approximately between amino acids 490-510; and other possible four alpha-helix parts. The remaining part of the C-terminal seems to be disordered which is in agreement with Partch et al [72] experimental results.

The UspA chaperone shows structural and also functional similarity to Cry2 N-terminal (alpha/beta domain), implying that Cry2 may have the chaperone activity. The cytosolic parts of Melanopsin, the C-tail of Melanopsin and the C-tail of Cry2 may take structure via binding to the chaperone part (N-terminal) of Cry2. The C-terminal of Cry may become alpha-helices via binding to the chaperone part of the Cry2.

The signal peptide sites found by different tools may sit in hCry2. If we consider the roles of these sites, we find out that they are parallel with the predicted roles of this protein:

- The autophosphorylation sites found can activate hCry2 in dark. By light activity; it may go and bind to Cell-Membrane by its Lipoprotein Attachment Region. Afterwards it can dock to Melanopsin, and start signal transduction.

According to Hex 4.5 results alpha/beta domain of hCry2 makes homodimer and also the C-tail binds to the alpha/beta region. The homodimer interaction via N-terminal most probably makes a conformational change in the protein. When homodimerization will occur, hCry2 should open up; meaning that its alpha domain should get apart from its alpha/beta domain. In dark phase, the homodimer can break up; the C-tail of the protein can reach to alpha/beta domain. Due to the chaperone activity of alpha/beta domain, the C-tail may change its structure. The nuclear localization signal site may become an alpha-helix and make the protein localized in the nucleus.

The homodimerization and open conformation also make the lipid-attachment site exposed to aqueous environment and the protein can bind to cell-membrane.

Possible Model for hCry2 Activity: According to Harada's work [50]; the nuclear-localization signal position of hCry2; the possible chaperone activity of the protein; possible switch between C-tail structure of model #3 and model #6; Hex 4.5 results for interaction between C-tail and UspA domain of hCry2 and homodimer interaction of hCry2.

It is possible that in light-phase hCry2 makes homodimer via staying at the open conformation (Figure 7.1). In open-conformation the C-tail cannot get in touch with the UspA domain of hCry2; consequently its C-tail structure is of model # 3. hCry2 is located in the cytoplasm and performs signal transduction via interacting with Melanopsin.

In dark-phase, homodimer breaks, hCry2 takes the closed conformation (Figure 7.2). The C-tail can contact to the UspA domain; so C-tail can take the structure of model # 6. Consequently, the nuclear-localization signal portion of the protein takes alpha-helix structure and protein can go into the nucleus after which protein can be phosphorylated and degraded.

The lipid-attachment region between residues 411-421 was found by *MotifScan* [26] tool. This region is covered (Figure 7.3) by a part of UspA domain, by residues 163-180. When hCry2 is in its the possible open conformation, this coverage disappears and lipid-attachment region becomes exposed to aqueous environment (Figure 7.4). Since this region is hydrophobic, it searches for hydrophobic region to bind to. Consequently, the open conformation makes it possible for hCry2 to bind to membrane.

The open-close conformation; homodimer formation and C-tail binding to N-terminal pathway are summarized at Figure 7.5. In the possible open conformation the alpha and alpha/beta domains separate from each other.

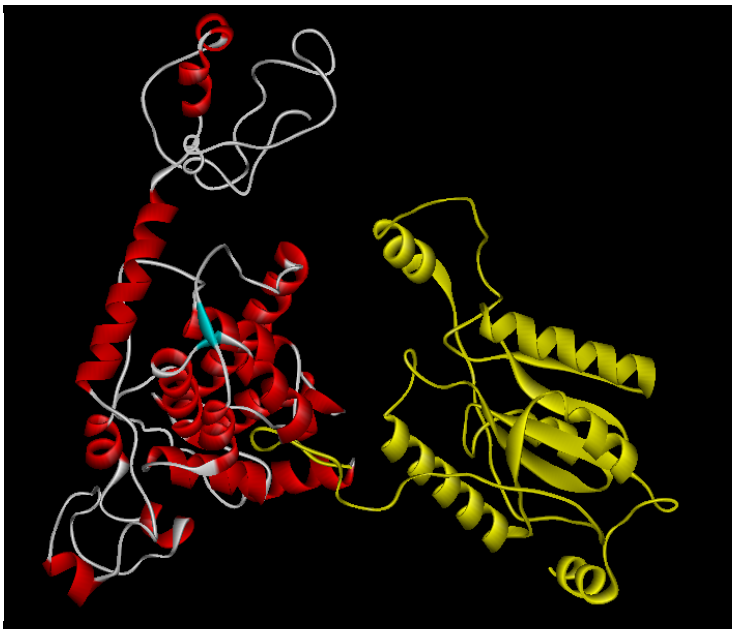


Figure 7.1: The open conformation of hCry2.

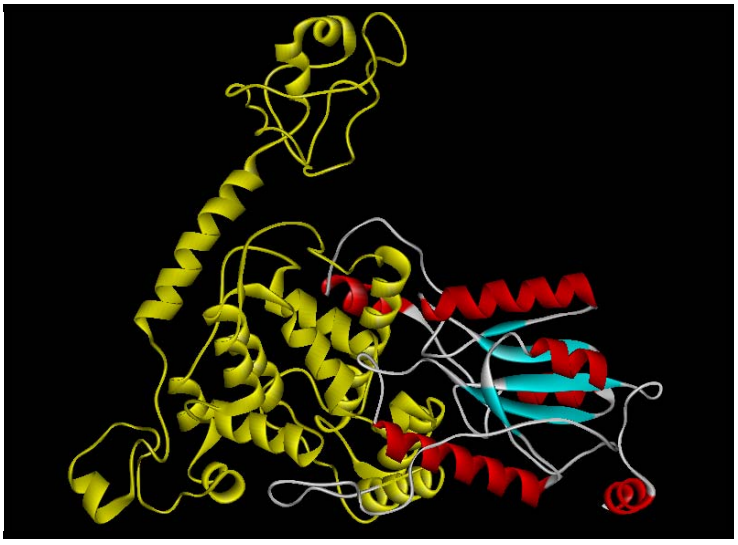


Figure 7.2: The closed conformation of hCry2.

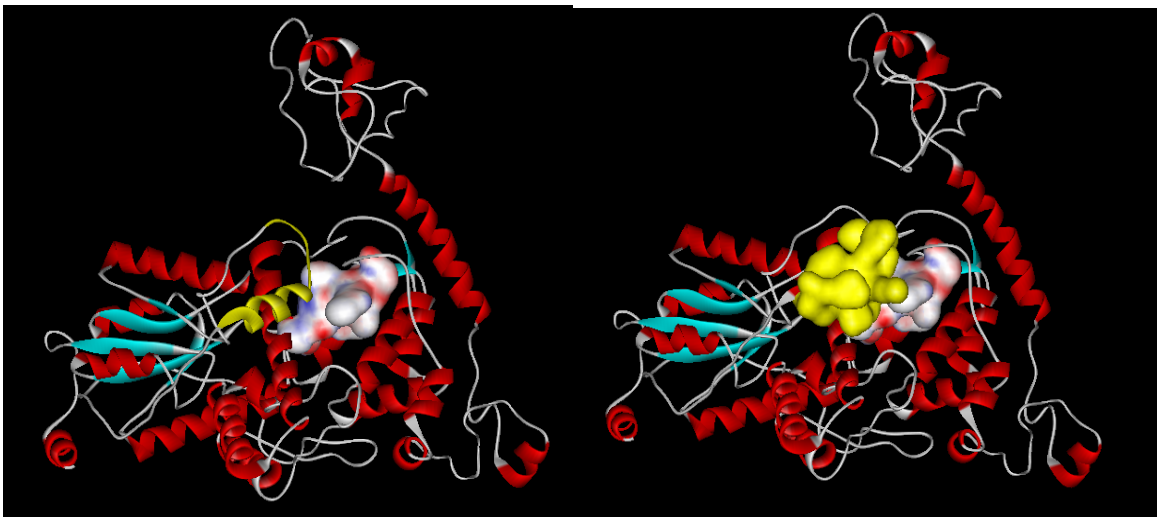


Figure 7.3: The lipid-attachment region is shown as surface; the region covering it (163-180 residues) is shown with yellow color at the left and as yellow surface at the right. The protein is in its closed-conformation.

Figure 7.4: The open conformation of hCry2.
The lipid-attachment region is exposed, it is shown as surface.
The covering region (163-180 residues) is shown with the yellow circle.

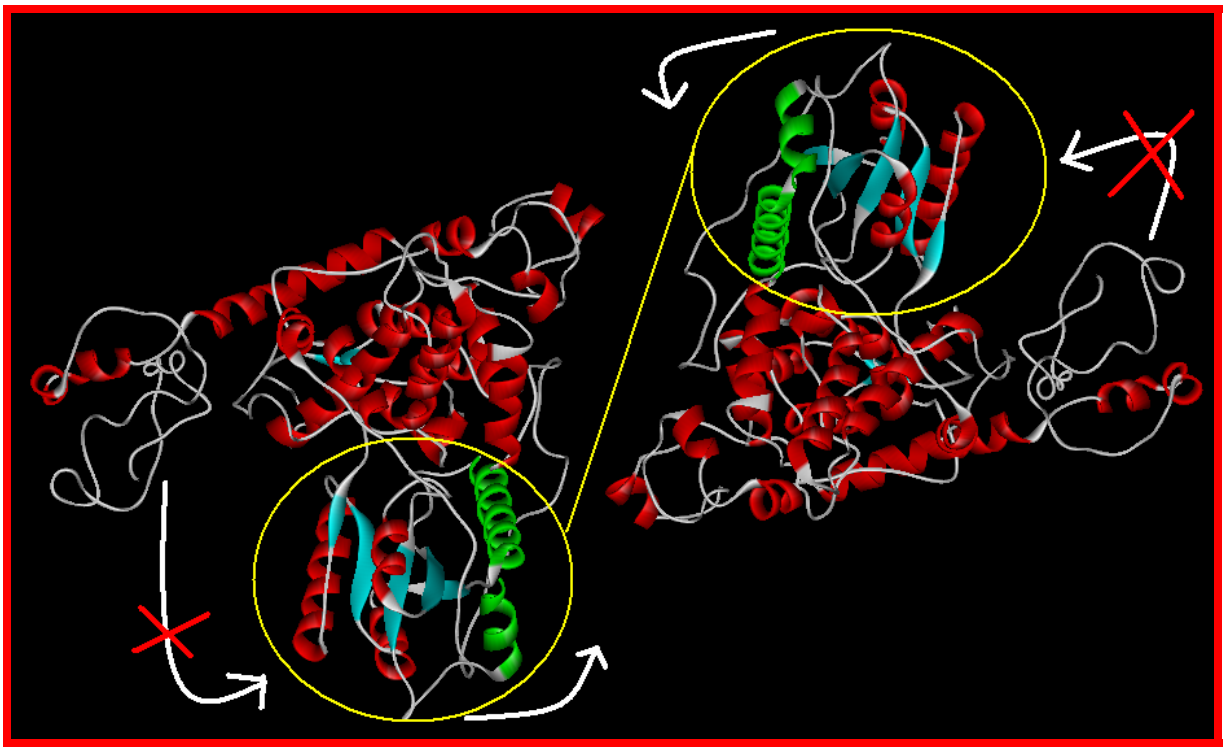


Figure 7.5: The open-close ability of hCry2; and possible homodimer formation. Homodimer and open-conformation formation are shown with white arrows; in open form it is not possible for C-tail to bind to UspAdomain and take structure. The parts that will form a homodimer are labeled with yellow circles.

All of these results lead possible pathways for hCry2-Melanopsin interaction; in light phase hCry2 makes homodimer and takes the open-conformation. This open conformation makes C-tail of the protein impossible to get in contact with chaperone portion (alpha/beta domain); consequently the C-tail has 2 little alpha-helices between residues 514-528; the other parts of C-tail are unstructured. In this open-conformation, the lipid-attachment site becomes exposed to aqueous environment and it can bind to cell-membrane. The cytosolic

loops of Melanopsin are now near to hCry2; so the interaction between them can take place. Consequently, signal transduction starts.

In dark phase, the homodimer breaks up, the interaction between Melanopsin and hCry2 also breaks up. The C-tail of hCry2 can reach to its alpha/beta domain. Due to the chaperone activity, C-tail may take structure; 2 alpha-helices at residues 559-565 and 575-583. The 559-565 regions have the nuclear localization signal; so the protein can localized to nucleus now; playing role of transcription factor and afterwards it can be degraded by phosphorylation upon residues 575-583.

The hypothesized functions and interactions of hCry2; the signaling cascade of Melanospin-Cry2 should be examined experimentally as a future work. When the computational results are verified by experimental methods; the biological pathway of mammalians will become clearer.

3D Homology modeling Tools Library

- Geno3D
- ESyPred
- CPHmodels
- SAM-T02
- HMMSTR
- LIBELLUA
- SDSC1
- 3D-PSSM
- PredictProtein
- Consensus
- PSIPRED
- EBI-SSM

Superimposition Tools Library

- Combinatorial Extension (CE) Method
- EMBL-Dali
- PINTS
- ProSup
- COMPARER
- SARF2
- FATCAT

Superimposition of real and predicted 3D structures of *A. thaliana* Cry2 by CE and FatCat methods:

Program Name	EsyPred Model
rmsd	0,20
Z-Score	8,30
Alignment	483
Gap	0

Program Name	CPH Model
rmsd	0,50
Z-Score	8,10
Alignment	388
Gap	0

Program Name	3Dpssm Model
rmsd	2,05
Z-Score	none
Alignment	379,00
Gap	14,00

Program Name	Consensus Model
rmsd	0
Z-Score	8,20
Alignment	388
Gap	0

Program Name	Swiss Model
rmsd	0,40
Z-Score	8,10
Alignment	388
Gap	0

Program Name	Geno3D Model
rmsd	1,65
Z-Score	none
Alignment	387

```
# this script is wirtten for analysis of output files of hex 4.5 program #

import sys
from math import *

## sometimes the length of "dockxxxx.pdb" and "charge.pdb" files are not equal, in order to  equilibrate them the number of charges and the number of dock runs should be compared ##

try:
   fc1=open("charge1.pdb","r")
except IOError:
   print >> sys.stderr, "Cannot open File!"
   sys.exit(1)

try:
   fc2=open("charge2.pdb","r")
except IOError:
   print >> sys.stderr, "Cannot open File!"
   sys.exit(1)

charge1lines=[]
charge1lines=fc1.readlines()

charge2lines=[]
charge2lines=fc2.readlines()

try :
   fen= open( 'electrostatic_energy.txt', 'w' )
except IOError :
   print >> sys.stderr, "Cannot open File!"
   sys.exit(1)

try :
   fen2= open( 'vanderwaals_energy.txt', 'w' )
except IOError :
   print >> sys.stderr, "Cannot open File!"
   sys.exit(1)

el_average=0.0
vdw_average=0.0
minimel=5000000.0
minimvdw=5000000.0

# to be able to read all file names in .pdb format #
for i in range(1,101):
   lines=[]
   c1=[]
```

```
c2=[]
if i==1 or 1<i<10:
   a=str(i)
   b="dock000"+a+".pdb"
if i==10 or 10<i<100:
   a=str(i)
   b="dock00"+a+".pdb"
if i==100:
   a=str(i)
   b="dock0"+a+".pdb"

try:
   file=open(b,"r")
except IOError:
   print >> sys.stderr, "Cannot open File!"
   sys.exit(1)

try :
   f1= open( 'pos1.txt', 'w' )
except IOError :
   print >> sys.stderr, "Cannot open File!"
   sys.exit(1)

try :
   f2= open( 'pos2.txt', 'w' )
except IOError :
   print >> sys.stderr, "Cannot open File!"
   sys.exit(1)

# to read lines of a file #
lines=file.readlines()

# to take the coordinates of atoms #
sayi=1000000
for j in range(8,len(lines)):
   asil=[]

   # to make ['a 55 r4t \n'] --> ['a', '55', 'r4t'] #
   asil=lines[j].split()
        # the lines starting with TER or REMARK should not be read #
        # after TER the position of second proteins atoms start #

   if asil[0]=="TER" or asil[0]=="REMARK":
      sayi=j

   if asil[0]=='ATOM':
      if asil[4] == "A" or asil[4] == "B" or asil[4] == "C":
```

```
            del asil[4]

        # write first atom's positions to a file, f1 #
        if(j<sayi):
            c1=charge1lines[i-8].split()
# to correct the charge lines #
            if asil[2]!=c1[2]:
                del asil
            else:
                f1.write(asil[5])
                f1.write(" ")
                f1.write(asil[6])
                f1.write(" ")
                f1.write(asil[7])
                f1.write("\n")
        if(j==sayi):
            print "Termination of first atom \n"

        # write second atom's positions to a file, f2 #
        if(j>sayi):
            c2=charge2lines[i-8].split()
            if asil[2]!=c2[2]:
                del asil
            else:
                f2.write(asil[5])
                f2.write(" ")
                f2.write(asil[6])
                f2.write(" ")
                f2.write(asil[7])
                f2.write("\n")

    f1.close()
    f2.close()

########################################
#charge & position calculation starts  ###
########################################


    try :
        f1= open( 'pos1.txt', 'r' )
    except IOError :
        print >> sys.stderr, "Cannot open File!"
        sys.exit(1)

    try :
        f2= open( 'pos2.txt', 'r' )
```

```
except IOError :
   print >> sys.stderr, "Cannot open File!"
   sys.exit(1)

try :
   f11= open( 'charge1.txt', 'r' )
except IOError :
   print >> sys.stderr, "Cannot open File!"
   sys.exit(1)
try :
   f22= open( 'charge2.txt', 'r' )
except IOError :
   print >> sys.stderr, "Cannot open File!"
   sys.exit(1)


### positions of fisrt atom ###
line1=f1.readlines()
### charges of first atom ###
line11=f11.readlines()
### position for second atom ###
line2=f2.readlines()
### charges of second atom ###
line22=f22.readlines()


el_energy=0.0
vdw_energy=0.0


for k in range(len(line1)):
   p1=[]
   c1=[]
   # to make ['a 55 r4t \n'] --> ['a', '55', 'r4t'] #
   p1=line1[k].split()
   c1=line11[k].split()
   for m in range(len(line2)):
      p2=[]
      c2=[]
      p2=line2[m].split()
      c2=line22[m].split()

      cons=sqrt(((float(p1[0])-float(p2[0]))**2) + ((float(p1[1])-float(p2[1]))**2) + ((float(p1[2])-
float(p2[2]))**2))
      if cons<=7:
      # calculate electrostatic energy #
         el_energy += float(c1[0])*float(c2[0])/cons
```

```
        # calculate Van der Waals energy #
           aaa=cons**12
           bbb=cons**6
           vdw_energy += 4*((1/aaa)-(1/bbb))

   el_average+=el_energy

   vdw_average+=vdw_energy

   if el_energy <= minimel:
      minimel = el_energy

   if vdw_energy <= minimvdw:
      minimvdw = vdw_energy

   fen.write("%d th " %i)
   fen.write("electrostatic energy is: %f " %el_energy)
   fen.write("\n")

   fen2.write("%d th " %i)
   fen2.write("Van der Waals energy is: %f " %vdw_energy)
   fen2.write("\n")


   f1.close()
   f2.close()
   f11.close()
   f22.close()

el_average=el_average/100.0
vdw_average=vdw_average/100.0

fen.write("\n")
fen.write("The AVERAGE ELECTROSTATIC ENERGY of 100 docks is; %f \n\n" %el_average)
fen.write("The MINIMUM ENERGY is; %f \n" %(minimel))


fen2.write("\n")
fen2.write("The AVERAGE VAN DER WAALS ENERGY of 100 docks is; %f \n\n" %vdw_average)
fen2.write("The MINIMUM ENERGY is; %f \n" %(minimvdw))

fen.close()
fen2.close()
fc1.close()
fc2.close()
```

Mini-Prep Protocol

1.5 ml sample should be put into 2 ml eppendorph tube

Centrifuge 20 seconds

Discard supernatant

Add 150 µl GTE and Vortex

Add 200 µl NaOH-SDS and Mix gently

Tube should be kept on ice for 5 minutes

Add 150 µl KAc and Mix gently

Tube should be kept on ice for 5 minutes

Centrifuge at maximum speed for 15 minutes

Discard pellet

Add 800 µl 95 % EtOH

Tube should be kept at room-temperature for 2 minutes

Centrifuge at maximum speed for 15 minutes

Discard supernatant

Add 400 µl 70 % EtOH

Centrifuge at maximum speed for 5 minutes

Discard supernatant

Dry the pellet

Add 50 µl TE buffer

Maxi-Prep

For 100 sample;

Centrifuge at 5000 rpm for 5 minutes

Discard supernatant

Add 4 ml GTE-RNase and Vortex

Tube should be kept at room-temperature for 10 minutes

Add 5 ml NaOH-SDS and Mix gently

Tube should be kept on ice for 5 minutes

Add 2 ml KAc and Mix gently

Tube should be kept on ice for 5 minutes

Centrifuge at 5000 rpm for 10 minutes

Discard pellet

Add 1:1 Chloroform and Vortex

Centrifuge at 5000 rpm for 10 minutes

Take the upper phase

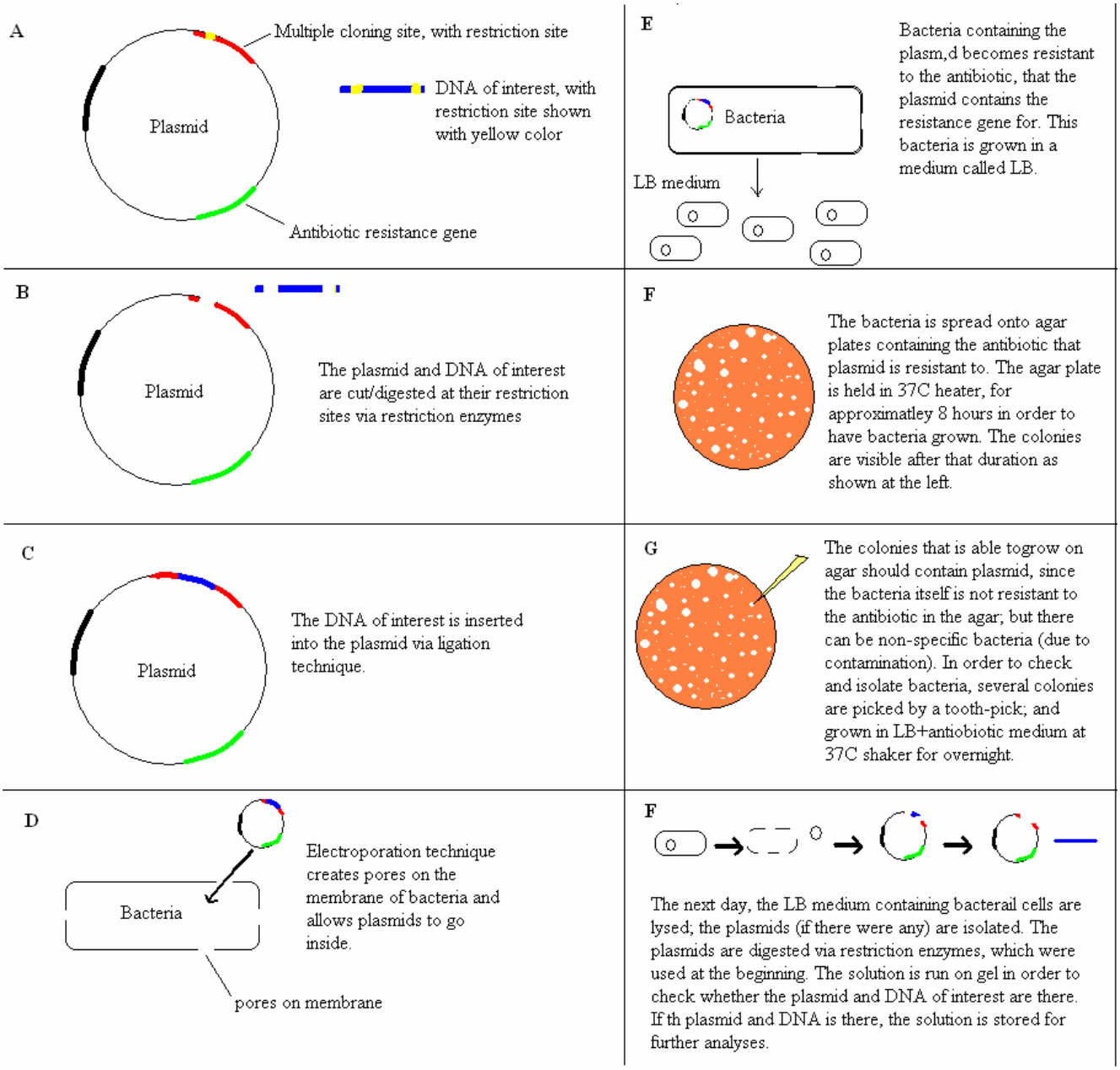
Add 0.7 Volume Isopropanol

Tube should be kept at room temperature for 5 minutes

Centrifuge at 13000 rpm for 15 minutes

Discard supernatant

Dry the pellet and Add 300-600 µl TE buffer according to the size of pellet


A
Multiple cloning site, with restriction site
DNA of interest, with restriction site shown with yellow color
Plasmid
Antibiotic resistance gene
B
Plasmid
The plasmid and DNA of interest are cut/digested at their restriction sites via restriction enzymes
C
Plasmid
The DNA of interest is inserted into the plasmid via ligation technique.
D
Bacteria
Electroporation technique creates pores on the membrane of bacteria and allows plasmids to go inside.
pores on membrane
E
Bacteria
LB medium
Bacteria containing the plasm,d becomes resistant to the antibiotic, that the plasmid contains the resistance gene for. This bacteria is grown in a medium called LB.
F
The bacteria is spread onto agar plates containing the antibiotic that plasmid is resistant to. The agar plate is held in 37C heater, for approximatley 8 hours in order to have bacteria grown. The colonies are visible after that duration as shown at the left.
G
The colonies that is able togrow on agar should contain plasmid, since the bacteria itself is not resistant to the antibiotic in the agar; but there can be non-specific bacteria (due to contamination). In order to check and isolate bacteria, several colonies are picked by a tooth-pick; and grown in LB+antiobiotic medium at 37C shaker for overnight.
F
The next day, the LB medium containing bacterail cells are lysed; the plasmids (if there were any) are isolated. The plasmids are digested via restriction enzymes, which were used at the beginning. The solution is run on gel in order to check whether the plasmid and DNA of interest are there. If th plasmid and DNA is there, the solution is stored for further analyses.

The active sites on *E. coli* photolyase and *A. thaliana* Cry that differ than the active sites on human Cry2 are written in ***bold italic***.

Active Site Results for *E. coli* Photolyase:

InterProScan:
No signal peptide regions were found.

ScanProsite:

1. ***DNA photolyases***
2. ***N-glycosylation site***
3. Protein kinase C phosphorylation site
4. N-myristoylation site
5. Tyrosine sulfation site
6. Casein kinase II phosphorylation site

MotifScan:

1. 26 29 freq_pat:ASN_GLYCOSYLATION [?]
2. 335 338 freq_pat:ASN_GLYCOSYLATION [?]
3. 389 392 freq_pat:CK2_PHOSPHO_SITE [?]
4. 404 407 freq_pat:CK2_PHOSPHO_SITE [?]
5. 430 433 freq_pat:CK2_PHOSPHO_SITE [?]
6. 61 66 freq_pat:MYRISTYL [?]
7. 138 143 freq_pat:MYRISTYL [?]
8. 165 170 freq_pat:MYRISTYL [?]
9. 266 271 freq_pat:MYRISTYL [?]
10. 28 30 freq_pat:PKC_PHOSPHO_SITE [?]
11. 38 40 freq_pat:PKC_PHOSPHO_SITE [?]
12. 48 50 freq_pat:PKC_PHOSPHO_SITE [?]
13. 120 122 freq_pat:PKC_PHOSPHO_SITE [?]
14. 180 182 freq_pat:PKC_PHOSPHO_SITE [?]
15. 235 237 freq_pat:PKC_PHOSPHO_SITE [?]
16. 248 250 freq_pat:PKC_PHOSPHO_SITE [?]
17. 302 304 freq_pat:PKC_PHOSPHO_SITE [?]
18. ***322 334 pat:DNA_PHOTOLYASES_1_1 [!]***
19. ***342 361 pat:DNA_PHOTOLYASES_1_2 [!]***
20. 2 177 pfam_fs:DNA_photolyase [!]
21. 202 469 pfam_fs:FAD_binding_7 [!]
22. 2 177 pfam_ls:DNA_photolyase [!]

23. 202 469 pfam_ls:FAD_binding_7 [!]

Active Site Results for *A. thaliana* Cry:

InterProScan:
No signal peptides were found.

ScanProsite:

1. ***DNA photolyases class 1 signature 1***
2. ***DNA photolyases class 1 signature 2***
3. Protein kinase C phosphorylation site
4. N-myristoylation site
5. Casein kinase II phosphorylation site
6. ***N-glycosylation site***
7. Tyrosine sulfation site
8. ***Tyrosine kinase phosphorylation site***
9. ***Bipartite nuclear targeting sequence***

MotifScan:

1. ***242 245 freq_pat:ASN_GLYCOSYLATION [?]***
2. ***478 481 freq_pat:ASN_GLYCOSYLATION [?]***
3. 114 117 freq_pat:CK2_PHOSPHO_SITE [?]
4. 188 191 freq_pat:CK2_PHOSPHO_SITE [?]
5. 215 218 freq_pat:CK2_PHOSPHO_SITE [?]
6. 278 281 freq_pat:CK2_PHOSPHO_SITE [?]
7. 384 387 freq_pat:CK2_PHOSPHO_SITE [?]
8. 403 406 freq_pat:CK2_PHOSPHO_SITE [?]
9. 454 457 freq_pat:CK2_PHOSPHO_SITE [?]
10. 465 468 freq_pat:CK2_PHOSPHO_SITE [?]
11. 506 509 freq_pat:CK2_PHOSPHO_SITE [?]
12. 71 76 freq_pat:MYRISTYL [?]
13. 241 246 freq_pat:MYRISTYL [?]
14. 420 425 freq_pat:MYRISTYL [?]
15. 459 464 freq_pat:MYRISTYL [?]
16. 534 539 freq_pat:MYRISTYL [?]
17. 66 68 freq_pat:PKC_PHOSPHO_SITE [?]
18. 114 116 freq_pat:PKC_PHOSPHO_SITE [?]
19. 236 238 freq_pat:PKC_PHOSPHO_SITE [?]
20. 257 259 freq_pat:PKC_PHOSPHO_SITE [?]

21. 454 456 freq_pat:PKC_PHOSPHO_SITE [?]
22. 480 482 freq_pat:PKC_PHOSPHO_SITE [?]
23. 513 523 freq_pat:PROKAR_LIPOPROTEIN [!]
24. 336 348 pat:DNA_PHOTOLYASES_1_1 [!]
25. 356 375 pat:DNA_PHOTOLYASES_1_2 [!]
26. 1 44 prf:C1Q [?]
27. 506 527 prf:SER_RICH [?]
28. 481 498 pre:NLS_BP [?]
29. 5 175 pfam_fs:DNA_photolyase [!]
30. 211 420 pfam_fs:FAD_binding_7 [!]
31. 5 175 pfam_ls:DNA_photolyase [!]
32. 211 450 pfam_ls:FAD_binding_7 [!]

Active Site Results for Mouse Cry2:

InterProScan:
On the last 110 aminoacids a signal peptide region was found.

ScanProsite:
1. Protein kinase C phosphorylation site
2. N-myristoylation site
3. Casein kinase II phosphorylation site
4. Tyrosine sulfation site
5. cAMP- and cGMP-dependent protein kinase phosphorylation site
6. ATP/GTP-binding site motif A (P-loop)
7. Cell attachment sequence
8. Amidation site

MotifScan:
All of the active sites are the same as human Cry2 active sites found by MotifScan.

The WHATIF Structural Evaluation Results:

The "Structural Z-scores" should have zero or positive value. The "RMS Z-scores" should be close to 1. When Robetta model # 3 WHATIF results are compared to the real structure of some proteins' WHATIF results, it can be easily seen that Robetta 3rd model has really a stable structure.

Robetta model #3:

Structure Z-scores, positive is better than average:
1st generation packing quality : -1.715
2nd generation packing quality : -2.789
Ramachandran plot appearance : -0.193
chi-1/chi-2 rotamer normality: 0.109
Backbone conformation : -0.833

RMS Z-scores should be close to 1.0:
Bond lengths : 0.809
Bond angles : 0.879
Omega angle restraints : 0.441 (tight)
Side chain planarity : 1.792
Improper dihedral distribution: 5.257 (loose)
Inside/Outside distribution : 1.084

A thaliana CRY2 real structure:

Structure Z-scores, positive is better than average:
1st generation packing quality: -0.886
2nd generation packing quality: -1.336
Ramachandran plot appearance : -0.300
chi-1/chi-2 rotamer normality: -0.320
Backbone conformation : 0.406

RMS Z-scores should be close to 1.0:
Bond lengths : 0.410 (tight)
Bond angles : 0.721
Omega angle restraints : 0.256 (tight)
Side chain planarity : 0.310 (tight)
Improper dihedral distribution: 0.494

Inside/Outside distribution : 1.050

UspA real structure:

Structure Z-scores, positive is better than average:
1st generation packing quality: 0.710
2nd generation packing quality: -0.482
Ramachandran plot appearance : 0.008
chi-1/chi-2 rotamer normality: -0.246
Backbone conformation : 1.266

RMS Z-scores should be close to 1.0:
Bond lengths : 0.415 (tight)
Bond angles : 0.678
Omega angle restraints : 0.323 (tight)
Side chain planarity : 0.302 (tight)
Inside/Outside distribution : 1.008

	1	2	3	4	5	6	7	8	9	10	Rosetta	PredictProtein	Rosetta HMMER
1	C	C	C	C	C	C	C	C	C	C	H	C 1.00	C
2	C	C	C	C	C	C	C	C	C	C	H	H 0.30	C
3	C	C	**H**	**H**	C	C	C	C	C	C	**H**	**H 0.50**	**H**
4	C	C	**H**	**H**	**H**	C	**H**	C	C	C	**H**	**H 0.90**	**H**
5	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
6	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
7	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
8	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
9	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.70**	**H**
10	**H**	**H**	**H**	**H**	C	**H**	C	**H**	**H**	**H**	**H**	**H 0.50**	**H**
11	**H**	**H**	**H**	C	C	**H**	C	**H**	**H**	**H**	**H**	**H 0.40**	**H**
12	H	H	H	C	C	H	C	H	H	H	C	C 0.60	C
13	H	H	H	C	C	H	C	H	H	H	C	C 0.80	C
14	H	H	H	C	C	H	C	H	H	H	C	C 0.90	C
15	H	H	H	C	C	H	C	H	H	H	C	C 0.90	C
16	H	H	H	C	C	H	C	H	H	H	C	C 1.00	C
17	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
18	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
19	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
20	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
21	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
22	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
23	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	C	C 0.60	C
24	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	C 0.20	E
25	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.50**	**E**
26	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.40**	**E**
27	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.30**	**E**
28	C	C	C	C	C	C	C	C	C	C	E	E 0.20	E
29	C	C	C	C	C	C	C	C	C	C	C	C 0.30	C
30	C	C	C	C	C	C	C	C	C	C	C	C 0.10	C
31	C	C	C	C	C	C	C	C	C	C	C	C 0.30	C
32	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
33	C	C	C	C	C	C	C	C	C	C	C	C 0.50	C
34	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
35	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
36	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
37	C	C	C	C	C	C	C	C	C	C	C	C 0.30	C
38	C	C	C	C	C	C	C	C	C	C	C	H 0.90	H
39	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C	**H 1.00**	**H**
40	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
41	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
42	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
43	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
44	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
45	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
46	H	H	H	H	H	H	H	H	H	H	C	H 0.40	C
47	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
48	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
49	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
50	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.30**	**E**
51	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.90**	**E**
52	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 1.00**	**E**
53	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 1.00**	**E**
54	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 1.00**	**E**

55	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.80**	**E**
56	E	E	E	E	E	E	E	E	E	E	C	E 0.50	E
57	C	C	C	C	C	C	C	C	C	C	C	C 0.30	C
58	C	C	C	C	C	C	C	C	C	C	C	C 0.60	H
59	C	C	C	C	C	C	C	C	C	C	C	C 0.60	H
60	C	C	C	C	C	C	C	C	C	C	C	C 0.50	H
61	C	C	C	C	C	C	C	C	C	C	H	C 0.40	H
62	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
63	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
64	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
65	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
66	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
67	C	C	H	C	C	C	C	C	C	C	C	C 0.80	C
68	H	H	H	C	C	H	C	H	H	H	H	H 0.40	H
69	H	H	H	C	H	H	H	H	H	H	H	H 0.50	H
70	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
71	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
72	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
73	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
74	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
75	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
76	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
77	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
78	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
79	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
80	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
81	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
82	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
83	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
84	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
85	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
86	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
87	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.70**	**H**
88	H	H	H	H	H	H	H	H	H	H	H	C 0.30	C
89	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
90	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
91	C	C	C	C	C	C	C	C	C	C	E	C 0.40	C
92	C	C	C	C	C	C	C	C	C	C	E	E 0.90	E
93	C	C	C	C	C	C	C	C	C	C	E	E 1.00	E
94	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.90**	**E**
95	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.70**	**E**
96	E	E	E	E	E	E	E	E	E	E	E	C 0.50	E
97	C	C	C	C	C	C	C	C	C	C	C	C 1.00	E
98	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
99	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
100	C	C	C	C	C	C	C	C	C	C	C	H 0.80	H
101	C	C	C	C	C	C	C	C	C	C	H	H 1.00	H
102	C	C	C	C	C	C	C	C	C	C	H	H 1.00	H
103	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
104	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
105	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
106	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
107	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
108	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
109	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.70**	**H**
110	H	H	H	H	H	H	H	H	H	H	H	C 0.20	H

111	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
112	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
113	C	C	C	C	C	C	C	C	C	C	C	C 0.40	C
114	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.40**	**E**
115	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.50**	**E**
116	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.70**	**E**
117	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.40**	**E**
118	E	E	E	E	E	E	E	E	E	E	E	E 0.20	C
119	C	C	C	C	C	C	C	C	C	C	C	C 0.40	C
120	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
121	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
122	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
123	C	C	C	C	C	C	C	C	C	C	H	C 0.60	C
124	H	H	H	H	H	H	H	H	H	H	H	C 0.60	C
125	H	H	H	H	H	H	H	H	H	H	H	C 0.40	H
126	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.30**	**H**
127	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.30**	**H**
128	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.50**	**H**
129	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
130	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
131	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
132	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
133	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
134	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
135	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
136	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
137	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.70**	**H**
138	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.40**	**H**
139	H	H	H	H	H	H	H	H	H	H	C	C 0.80	C
140	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
141	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
142	C	C	C	C	C	C	C	C	C	C	E	E 0.60	E
143	C	C	C	C	C	C	C	C	C	C	E	E 0.80	E
144	C	C	C	C	C	C	C	C	C	C	E	E 0.70	E
145	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.50**	**E**
146	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E 0.20**	**E**
147	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
148	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
149	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
150	C	C	C	C	C	C	C	C	C	C	C	C 0.40	E
151	C	C	C	C	C	C	C	C	C	C	C	C 0.40	E
152	C	C	C	C	C	C	C	C	C	C	C	C 0.60	E
153	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
154	C	C	C	C	C	C	C	C	C	C	C	C 0.90	H
155	C	C	C	C	C	C	C	C	C	C	C	C 0.60	H
156	C	C	C	C	C	C	C	C	C	C	C	C 0.50	H
157	C	C	C	C	C	C	C	C	C	C	E	C 0.50	C
158	C	C	C	C	C	C	C	C	C	C	E	C 0.50	C
159	C	C	C	C	C	C	C	C	C	C	E	C 0.60	C
160	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
161	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
162	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
163	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
164	C	C	C	C	C	C	C	C	C	C	E	C 1.00	C
165	C	C	C	C	C	C	C	C	C	C	E	C 0.90	C
166	C	C	C	C	C	C	C	C	C	C	C	C 0.60	E

167	C	C	C	C	C	C	C	C	C	C	C	C 0.40	E
168	C	C	C	C	C	C	C	C	C	C	C	C 0.30	E
169	C	C	**H**	C	**H**	C	**H**	C	C	C	**H**	C 0.20	C
170	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.50**	**H**
171	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.70**	**H**
172	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
173	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
174	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
175	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
176	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
177	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
178	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.70**	**H**
179	H	H	C	H	C	H	C	H	H	H	H	H 0.30	C
180	C	C	C	C	C	C	C	C	C	C	H	C 0.60	C
181	C	C	C	C	C	C	C	C	C	C	H	C 0.80	C
182	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
183	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
184	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
185	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
186	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
187	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
188	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
189	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
190	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
191	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
192	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
193	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
194	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
195	C	C	C	C	C	C	C	C	C	C	E	C 0.80	C
196	C	C	C	C	C	C	C	C	C	C	E	C 0.80	C
197	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
198	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
199	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
200	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
201	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
202	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
203	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
204	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
205	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
206	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
207	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
208	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
209	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
210	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
211	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
212	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
213	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
214	C	C	C	C	C	C	C	C	C	C	C	C 0.60	H
215	C	C	C	C	C	C	C	C	C	C	C	C 0.50	H
216	C	C	C	C	C	C	C	C	C	C	C	C 0.50	H
217	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
218	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
219	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
220	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
221	C	C	C	C	C	C	C	C	C	C	C	C 0.80	H
222	C	C	C	C	C	C	C	C	C	C	C	C 0.80	H

223	C	C	C	C	C	C	C	C	C	C	C	C 0.80	H
224	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
225	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
226	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
227	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
228	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
229	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
230	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
231	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
232	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
233	C	C	C	C	C	C	C	C	C	C	H	H 0.20	C
234	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
235	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
236	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
237	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
238	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
239	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
240	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
241	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
242	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
243	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
244	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
245	C	C	C	C	C	C	C	C	C	C	H	H 0.90	H
246	C	C	C	C	C	C	C	C	C	C	H	H 0.70	H
247	C	C	C	C	C	C	C	C	C	C	H	H 0.30	C
248	C	C	C	C	C	C	C	C	C	C	H	C 0.30	C
249	C	C	C	C	C	C	C	C	C	C	H	C 0.30	C
250	C	C	C	C	C	C	C	C	C	C	H	H 0.20	C
251	C	C	C	C	C	C	C	C	C	C	H	H 0.20	C
252	C	C	C	C	C	C	C	C	C	C	H	H 0.30	C
253	C	C	C	C	C	C	C	C	C	C	H	C 0.20	H
254	C	C	H	C	C	C	C	C	C	C	H	C 0.50	H
255	C	C	H	C	C	C	C	C	C	C	H	C 0.60	H
256	C	C	H	C	C	C	C	C	C	C	H	C 0.60	C
257	C	C	H	C	C	C	C	C	C	C	C	C 0.80	C
258	C	C	H	C	C	C	C	C	C	C	C	C 0.90	C
259	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
260	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
261	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
262	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
263	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
264	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
265	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
266	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
267	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
268	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
269	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
270	C	C	C	C	C	C	C	C	C	C	H	C 0.50	C
271	**H**	**H**	**H**	**C**	**C**	**H**	**C**	**H**	**H**	**H**	**H**	C 0.30	**H**
272	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	E 0.30	**H**
273	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.20	**H**
274	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.20	**H**
275	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.20	**H**
276	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.20	C
277	C	C	C	C	C	C	C	C	C	C	C	C 0.40	C
278	C	C	C	C	C	C	C	C	C	C	C	C 0.50	C

279	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
280	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
281	C	C	C	C	C	C	C	C	C	C	H	H 0.50	H
282	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
283	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
284	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
285	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
286	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
287	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
288	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
289	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
290	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
291	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
292	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
293	**H**	**H**	**H**	**H**	**C**	**H**	**C**	**H**	**H**	**H**	**H**	**H 0.60**	**H**
294	H	H	C	H	C	H	C	H	H	H	C	H 0.40	H
295	H	H	C	H	C	H	C	H	H	H	H	C 0.30	H
296	H	H	C	C	C	H	C	H	H	H	C	C 0.70	C
297	H	H	C	C	C	H	C	H	H	H	C	C 0.90	C
298	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
299	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
300	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
301	C	C	H	C	H	C	H	C	C	C	C	C 0.80	C
302	C	C	H	C	H	C	H	C	C	C	C	H 0.30	H
303	C	C	**H**	C	**H**	C	**H**	C	C	C	**H**	**H 1.00**	**H**
304	C	C	**H**	C	**H**	C	**H**	C	C	C	**H**	**H 1.00**	**H**
305	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
306	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
307	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
308	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
309	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
310	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
311	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
312	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
313	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
314	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
315	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
316	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
317	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
318	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
319	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
320	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
321	H	H	H	H	H	H	H	H	H	H	H	H 0.30	H
322	H	H	H	H	H	H	H	H	H	H	C	C 0.80	C
323	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
324	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
325	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
326	C	C	C	C	C	C	C	C	C	C	C	C 0.60	H
327	C	C	C	C	C	C	C	C	C	C	C	C 0.60	H
328	C	C	C	C	C	C	C	C	C	C	C	C 0.70	H
329	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
330	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
331	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
332	C	C	C	C	C	C	C	C	C	C	C	C 0.50	H
333	C	C	C	C	C	C	C	C	C	C	C	C 0.40	H
334	C	C	C	C	C	C	C	C	C	C	C	C 0.40	H

335	C	C	C	C	C	C	C	C	C	C	C	C 0.50	H
336	C	C	C	C	C	C	C	C	C	C	C	C 0.40	H
337	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
338	C	C	C	C	C	C	C	C	C	C	E	C 0.70	C
339	C	C	C	E	C	C	C	C	C	C	C	C 0.80	C
340	E	E	C	E	C	E	C	E	E	E	C	C 0.60	C
341	E	E	C	C	C	E	C	E	E	E	C	C 0.50	C
342	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
343	C	C	C	C	C	C	C	C	C	C	H	H 0.40	H
344	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.60**	**H**
345	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
346	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
347	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
348	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
349	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.70**	**H**
350	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.50**	**H**
351	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.60**	C
352	C	C	C	C	C	C	C	C	C	C	H	C 0.40	C
353	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
354	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
355	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
356	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
357	C	C	C	C	C	C	C	C	C	C	C	H 0.30	C
358	H	H	H	H	H	H	H	H	H	H	C	H 0.40	H
359	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.40**	**H**
360	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
361	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
362	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
363	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
364	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
365	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
366	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
367	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
368	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
369	C	C	C	C	C	C	C	C	C	C	C	H 0.60	H
370	C	C	C	C	C	C	C	C	C	C	C	C 0.20	C
371	C	C	C	C	C	C	C	C	C	C	C	H 0.10	C
372	C	C	C	C	C	C	C	C	C	C	C	H 0.30	H
373	C	C	C	C	C	C	C	C	C	C	C	H 0.30	H
374	C	C	C	C	C	C	C	C	C	C	H	H 0.30	H
375	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.70**	**H**
376	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
377	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
378	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
379	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
380	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
381	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
382	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
383	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
384	**H**	**H**	**H**	**H**	**C**	**H**	**C**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
385	**H**	**H**	**H**	**H**	**C**	**H**	**C**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
386	H	H	H	H	H	H	H	H	H	H	C	H 0.60	H
387	H	H	H	C	H	H	H	H	H	H	C	H 0.50	H
388	H	H	H	C	H	H	H	H	H	H	C	H 0.40	H
389	C	C	C	C	H	C	H	C	C	C	C	H 0.20	C
390	E	E	E	E	H	E	H	E	E	E	C	H 0.20	C

391	E	E	E	E	C	E	C	E	E	E	C	C 0.20	C
392	C	C	C	C	C	C	C	C	C	C	C	C 0.40	C
393	C	C	C	C	C	C	C	C	C	C	H	C 0.20	C
394	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.60**	**H**
395	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.60**	**H**
396	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
397	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
398	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
399	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
400	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
401	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
402	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
403	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
404	C	C	C	C	C	C	C	C	C	C	H	H 0.90	C
405	C	C	C	C	C	C	C	C	C	C	H	H 0.40	C
406	C	C	C	C	C	C	C	C	C	C	C	C 0.20	C
407	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
408	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
409	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
410	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.50	C
411	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.70	C
412	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.80	C
413	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.80	H
414	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.70	H
415	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.50	C
416	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	E 0.20	C
417	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	E 0.50	E
418	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	E 0.80	E
419	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	E 0.90	E
420	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	E 0.50	C
421	C	C	C	C	C	C	C	C	C	C	C	C 0.30	C
422	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
423	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
424	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
425	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
426	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
427	C	C	C	C	C	C	C	C	C	C	C	C 0.50	C
428	C	C	C	C	C	C	C	C	C	C	C	C 0.20	E
429	C	C	C	C	C	C	C	C	C	C	C	C 0.20	E
430	C	C	C	C	C	C	C	C	C	C	C	E 0.30	E
431	C	C	C	C	C	C	C	C	C	C	C	E 0.30	E
432	C	C	C	C	C	C	C	C	C	C	C	C 0.20	C
433	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
434	C	C	C	C	C	C	C	C	C	C	C	C 0.70	H
435	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.50	**H**
436	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.40	**H**
437	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.50	**H**
438	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.40	**H**
439	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.60	**H**
440	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.80	C
441	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C 0.90	C
442	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
443	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
444	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
445	C	C	C	C	C	C	C	C	C	C	C	C 0.20	H
446	C	C	C	C	C	C	C	C	C	C	C	C 0.10	H

447	H	H	H	H	H	H	H	H	H	H	C	C 0.20	H
448	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.60**	**H**
449	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
450	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
451	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
452	C	C	C	C	C	C	C	C	C	C	C	H 0.90	H
453	C	C	C	C	C	C	C	C	C	C	C	H 1.00	H
454	C	C	C	C	C	C	C	C	C	C	C	H 1.00	H
455	C	C	C	C	C	C	C	C	C	C	C	H 1.00	H
456	C	C	C	C	C	C	C	C	C	C	C	H 0.90	C
457	C	C	C	C	C	C	C	C	C	C	C	H 0.70	C
458	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
459	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
460	C	C	C	C	C	C	C	C	C	C	H	C 0.40	H
461	C	C	C	C	C	C	C	C	C	C	H	C 0.20	H
462	C	C	C	C	C	C	C	C	C	C	H	C 0.30	H
463	C	C	C	C	C	C	C	C	C	C	H	C 0.30	H
464	C	C	C	C	C	C	C	C	C	C	H	C 0.20	C
465	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
466	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
467	C	C	C	C	C	C	C	C	C	C	H	C 0.70	C
468	C	C	C	C	C	C	C	C	C	C	H	C 0.80	C
469	C	C	C	C	C	C	C	C	C	C	H	C 0.90	C
470	C	C	C	C	C	C	C	C	C	C	H	C 0.80	C
471	C	C	C	C	C	C	C	C	C	C	H	H 0.20	H
472	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.50**	**H**
473	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.50**	**H**
474	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.60**	**H**
475	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.60**	**H**
476	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.30**	**H**
477	H	H	H	H	H	H	H	H	H	H	C	C 0.60	C
478	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
479	C	C	C	C	C	C	C	C	C	C	C	C 0.50	C
480	C	C	C	C	C	C	C	C	C	C	E	C 0.40	C
481	C	C	C	C	C	C	C	C	C	C	E	C 0.50	C
482	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
483	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
484	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
485	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
486	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
487	C	C	C	C	C	C	C	C	C	C	C	C 0.60	H
488	C	C	C	C	C	C	C	C	C	C	C	C 0.50	H
489	C	C	C	C	C	C	C	C	C	C	C	C 0.40	H
490	C	C	C	C	C	C	C	C	C	C	C	C 0.50	C
491	C	C	C	C	C	C	C	C	C	C	C	C 0.30	C
492	C	C	C	C	C	C	C	C	C	C	H	C 0.20	H
493	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.50**	**H**
494	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
495	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.80**	**H**
496	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 0.90**	**H**
497	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
498	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
499	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
500	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
501	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
502	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**

503	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
504	**H**	**H**	**C**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
505	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
506	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
507	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
508	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
509	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
510	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H 1.00**	**H**
511	**H**	**H**	**H**	**H**	**H**	**H**	**H**	C	C	H	C	**H 0.90**	C
512	C	C	H	C	C	C	C	C	C	C	C	H 0.60	C
513	C	C	C	C	C	C	C	C	C	C	C	C 0.40	C
514	C	C	C	C	C	C	C	C	C	C	C	C 0.50	C
515	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
516	C	C	C	H	C	C	C	C	C	C	C	C 0.50	C
517	C	C	C	H	C	C	C	C	C	C	C	E 0.50	C
518	C	C	C	H	C	C	C	C	C	C	C	E 0.60	C
519	C	C	C	H	C	C	C	C	C	C	C	E 0.40	C
520	C	C	C	H	C	C	C	C	C	C	C	E 0.30	C
521	C	C	C	H	C	C	C	C	C	C	C	C 0.30	C
522	C	C	C	H	C	C	C	C	C	C	C	C 0.70	C
523	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
524	C	C	H	H	C	C	C	C	C	H	C	C 0.40	C
525	C	C	H	H	H	C	H	C	C	H	C	C 0.60	C
526	C	C	H	H	H	C	H	C	C	H	C	C 0.50	C
527	C	C	H	H	H	C	H	C	C	H	C	C 0.60	C
528	C	C	H	H	H	C	H	C	C	H	C	C 0.70	C
529	C	C	C	C	H	C	H	C	C	C	C	C 0.90	C
530	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
531	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
532	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
533	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
534	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
535	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
536	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
537	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
538	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
539	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
540	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
541	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
542	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
543	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
544	C	C	C	C	C	C	C	C	C	C	C	C 0.50	C
545	C	C	C	C	C	C	C	C	C	C	C	C 0.70	C
546	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
547	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
548	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
549	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
550	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
551	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
552	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
553	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
554	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
555	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
556	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
557	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
558	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C

559	C	C	C	C	C	C	C	C	C	C	C	H 0.30	C
560	C	C	C	C	C	C	C	C	C	C	C	**H 0.60**	C
561	**H**	**H**	C	C	C	**H**	C	C	C	**H**	C	**H 0.20**	C
562	**H**	**H**	C	C	C	**H**	C	C	C	**H**	C	C 0.30	C
563	H	H	C	C	C	H	C	C	C	H	C	C 0.20	C
564	H	H	C	C	C	H	C	C	C	H	C	C 0.30	C
565	H	H	C	C	C	H	C	C	C	H	C	C 0.60	C
566	H	H	C	C	C	C	C	C	C	C	C	C 0.90	C
567	H	H	C	C	C	C	C	C	C	C	C	C 0.70	C
568	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
569	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
570	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
571	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
572	C	C	C	C	C	**H**	C	C	C	C	C	**H 0.90**	C
573	C	C	C	C	C	**H**	C	C	C	C	C	**H 0.70**	C
574	C	C	C	C	C	**H**	C	C	C	C	**H 0.40**	**H 0.70**	C
575	C	C	C	C	C	**H**	C	C	C	C	**H 0.40**	**H 0.50**	E
576	C	C	C	C	C	**H**	C	C	C	C	**H 0.40**	C 0.30	E
577	C	C	C	C	C	**H**	C	C	C	C	**H 0.40**	C 0.20	E
578	**H**	**H**	C	C	**H**	**H**	**H**	C	C	C	**H 0.30**	C 0.30	E
579	H	H	C	C	H	H	H	C	C	C	C	C 0.20	E
580	H	H	C	C	H	H	H	C	C	C	C	C 0.20	E
581	H	H	C	C	H	H	H	C	C	C	C	C 0.50	C
582	H	H	C	C	H	C	H	C	C	C	C	C 0.60	C
583	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
584	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
585	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
586	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
587	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
588	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
589	C	C	C	C	C	C	C	C	C	C	C	C 0.80	C
590	C	C	C	C	C	C	C	C	C	C	C	C 0.60	C
591	C	C	C	C	C	C	C	C	C	C	C	C 0.90	C
592	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C
593	C	C	C	C	C	C	C	C	C	C	C	C 1.00	C

First column indicates the amino acid residue numbers. Columns 2-11 are the secondary structure of hCry2 according to "*Robetta* tertiary structure prediction" tool (i.e. column 2 is the prediction taken from 1st model). Last 3 columns are secondary structure predictions according to the "secondary structure prediction tools; *Rosetta*, *Rosetta-HMMER* and *Predict-Protein*.

The matches between all predictions are written in **bold**; the alpha-helice parts are shown with yellow color; whereas the beta-sheet parts are shown with blue color.

The model #3 (the fourth column) were modified according to those results.

Program Name	**CE**	**FatCat**
rmsd	4,20	3,14
Z-Score	2,80	-
Alignmet	48,00	113,00
Gap	24,00	30,00
Similarity	-	23,08%

UspA domain1 versus hCry2 superimposition

Program Name	**CE**	**FatCat**
rmsd	3,20	3,06
Z-Score	3,50	-
Alignmet	40,00	111,00
Gap	6,00	30,00
Similarity	-	20,57%

UspA domain2 versus hCry2 superimposition

Program Name	**CE**	**FatCat**
rmsd	2,50	2,76
Z-Score	3,70	-
Alignmet	36,00	109,00
Gap	6,00	25,00
Similarity	-	21,64%

UspA domain3 versus hCry2 superimposition

Program Name	**CE**	**FatCat**
rmsd	2,70	3,15
Z-Score	3,50	-
Alignmet	31,00	107,00
Gap	6,00	32,00
Similarity	-	25,18%

UspA domain4 versus hCry2 superimposition

Query is the sequence of hamster Melanopsin; subject is the sequence of human Melanopsin.

```
Query  1   MDSPPGPTAPPGLTQGPSFMASTTLHSHWNSTQK-VSTRAQLLAVSPTASGPEAAAWVPF  59
           M+ P GP  PP  TQ PS MA+    S W+S+Q  +S+  +L ++SPTA G  AAAWVP
Sbjct  1   MNPPSGPRVPPSPTQEPSCMATPAPPSWWDSSQSSISSLGRLPSISPTAPGTWAAAWVPL  60

Query  60
PTVDVPDHAHYXXXXXXXXXXXXXXXXXNLTVIYTFCRSRSLRTPANMLIINLAVSDFLMS  119
           PTVDVPDHAHY LGTVILLVGLTGMLGNLTVIYTFCRSRSLRTPANM IINLAVSDFLMS
Sbjct  61  PTVDVPDHAHYTLGTVILLVGLTGMLGNLTVIYTFCRSRSLRTPANMFIINLAVSDFLMS  120

Query  120  FTQAPVFFASSLYKKWLFGETGCEFYAFCGAVLGITSMITLTAIALDRYLVITRPLATIG  179
            FTQAPVFF SSLYK+WLFGETGCEFYAFCGA+ GI+SMITLTAIALDRYLVITRPLAT G
Sbjct  121  FTQAPVFFTSSLYKQWLFGETGCEFYAFCGALFGISSMITLTAIALDRYLVITRPLATFG  180

Query  180
MGSKRRTALVLLGIWLYALAWSLPPFFGWSAYVPEGLLTSCSWDYVTFTPQVRAYTMXXX
            + SKRR A VLLG+WLYALAWSLPPFFGWSAYVPEGLLTSCSWDY++FTP VRAYTMLL
Sbjct  181
VASKRRAAFVLLGVWLYALAWSLPPFFGWSAYVPEGLLTSCSWDYMSFTPAVRAYTMLLC  240

Query  240  XXXXXXXXXXXXXXYISIFRAIRETGR------ACEGWSESPQRRRQWHRLQSEWKMAKV  293
           CFVFFLPLL+II+CYI IFRAIRETGR      AC+G  ES  +R+   RLQSE KMAK+
Sbjct  241  CFVFFLPLLIIIYCYIFIFRAIRETGRALQTFGACKGNGESLWQRQ---RLQSECKMAKI  297

Query  294  ALIVILLFVLSWAPYSTVALVAFAGYSHILTPYMSSVPAVIAKASAIHNPIVYAITHPKY  353
             L+VILLFVLSWAPYS VALVAFAGY+H+LTPYMSSVPAVIAKASAIHNPI+YAITHPKY
Sbjct  298  MLLVILLFVLSWAPYSAVALVAFAGYAHVLTPYMSSVPAVIAKASAIHNPIIYAITHPKY  357

Query  354
RAAIAQHLPCLGVLLGVSSQRNRPSLSYXXXXXXXXXXXXXXXXXXXAPKRQESLGSESE  413
           R AIAQHLPCLGVLLGVS + +RP  SYRSTHRSTL+S +S+LSWIS  +RQESLGSESE
Sbjct  358  RVAIAQHLPCLGVLLGVSRRHSRPYPSYRSTHRSTLTSHTSNLSWISIRRRQESLGSESE  417

Query 414  VGWTDTEATAVWGAAQPASGQSSCGQNLEDGMVKAPSSPQ---------AKGQLPSLD  462
            VGWT  EA AVWGAAQ A+G+S  GQ LED   KAP  PQ         KG +PS D
Sbjct  418
       VGWTHMEAAAVWGAAQQANGRSLYGQGLEDLEAKAPPRPQGHEAETPGKTKGLIPSQD  475
```

1. One day prior to transfection, plate the cells. The cells then could logarithmically grow on the day of transfection (i.e. 60-80% confluent at the time of transfection).
2. Prepare two sets of 15 ml tubes.

   Into one set pipet: 0.5ml 2X HEBS

   Into the other set mix:

   DNA (5 ug in TE) to 450ul of TE

   2.5M Ca Cl_2.

3. Mix DNA and $CaCl_2$ by drawing up and down in a 1 ml pipet and add dropwise to the 2XHEBS. Bubbling should occur when the mixing is achieved.
4. Let the mixture stand at room temp for 8 minutes.
5. Add the DNA $CaPO_4$ coprecipitate dropwise to the surface of the media containing the cells. Swirl the plate gently to mix.
6. Incubate culture 7-11 hours, remove the $CaPO_4$ containing medium. Replace with normal medium.

The superimposition results of hCry2 model by *E. coli* DNA Photolyase protein. *E. coli* DNA Photolyase protein is 469 amino acids long protein. The match length of superimposition and the rmsd value are given.

Match Length	457
rmsd	3,03

The superimposition results of hCry2 model by *Synechocytsis* Cry protein. *Synechocytsis* Cry protein is 483 amino acids long protein. The match length of superimposition and the rmsd value are given.

Match Length	472
rmsd	1,37

1. *www.pdb.org*
2. *http://www.ebi.ac.uk/InterProScan/*
3. *http://myhits.isb-sib.ch/cgi-bin/motif_scan*
4. *http://au.expasy.org/tools/scanprosite/*
5. *http://www.fundp.ac.be/urbm/bioinfo/esypred/*
6. *http://structure.bu.edu/cgi-bin/consensus/consensus.cgi*
7. *http://robetta.bakerlab.org/*
8. *http://www.biochem.abdn.ac.uk/hex/*
9. *http://cl.sdsc.edu/ce.html*
10. *http://fatcat.burnham.org/fatcatpair.html*
11. *http://swift.cmbi.kun.nl/WIWWWI/*
12. *http://split.pmfst.hr/split/*
13. *http://jafa.burnham.org*
14. *http://psort.ims.u-tokyo.ac.jp*
15. *http://www.predictprotein.org*
16. *http://www.jens-meiler.de/turnpred.html*

BIBLIOGRAPHY

[1] I H Kavakli, A Sancar: Circadian Photoreception in Humans and Mice**,** Molecular Interventions (2002) 2:484
[2] M Ahmad, A R Cashmore, HY4 gene of Arabidopsis thaliana encodes a protein with characteristics of a blue-light photoreceptor, Nature (1993) 366:162
[3] A R Cashmore, J A Jarillo, Y J Wu, Cryptochromes: Blue light receptors for plants and animals (1999) 284:760
[4] C Lin, T Todo, The Cryptochromes, Genome Biology (2005) 6:220
[5] M Barinaga, Clock Photoreceptor Shared By Plants and Animals, Science (1998) 1628
[6] Y. Miyamoto and A. Sancar, Vitamin B2-based blue-light photoreceptors in the retinohypothalamic tract as the photoactive pigments for setting the circadian clock in mammals, Proc. Natl. Acad. Sci. USA (1998) 95:6097
[7] J C Hall, Cryptochrome: sensory reception, transduction and clock functions subserving circadian systems, Curr. Opin. Neurobiol. (2000) 10:456
[8] G T J Van der Horst, M Muijtjens, K Kobayashi, Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms, Nature (1999) 398:627
[9] M. H. Vitaterna, C. P. Selby, T. Todo, H. Niwa, C. Thompson, E. M. Fruechte, K. Hitomi, R. J. Thresher, T. Ishikawa, J. Miyazaki, J. S. Takahashi, and A. Sancar, Differential regulation of mammalian Period genes and circadian rhythmicity by cryptochromes 1 and 2, Proc. Natl. Acad. Sci. USA (1999) 96:12114
[10] R J Thresher, M H Vitaterna, Y Miyamoto, A Kazantsev, D S Hsu, C Petit, C P Selby, L Dawut, O Smithies, J S Takahashi, A Sancar, Role of Mouse Cryptochrome Blue-Light Photoreceptor in Circadian Photoresponses, Science (1998) 282:1490
[11] A. Sancar, Cryptochrome: The second photoactive pigment in the eye and its role in circadian photoreception, Annu Rev Biochem. (2000) 69:31
[12] T Todo, H Ryo, K Yamamoto, H Toh, T Inui, H Ayaki, T Nomura, M Ikenaga, Similarity Among the *Drosophila* (6-4)Photolyase, a Human Photolyase Homolog, and the DNA Photolyase-Blue-Light Photoreceptor Family, Science (1996) 272:109
[13] D S Hsu, X Zhao, S Zhao, A Kazantsev, R P Wang, T Todo, Y F Wei, A Sancar, Putative Human Blue-Light Photoreceptors hCRY1 and hCRY2 Are Flavoproteins**,** Biochemistry **(**1996**)** 35**:**13871
[14] L Provencio, G Jiang, W J DeGrip, W P Hayes, M D Rollag, Melanopsin: an opsin in melanophores, brain, and eye, Proc. Natl. Acad. Sci. U S A (1998) 95:340
[15] I Provencio, I R Rodriguez, G Jiang, W P Hayes, E . Moreira, M D Rollag, A Novel Human Opsin in the Inner Retina, J. Neurosci (2000) 20:600
[16] D M Berson, F A Dunn, M Takao, Phototransduction by Retinal Ganglion Cells That Set the Circadian Clock, Science (2002) 295:1070

[17] J S Takahashi, P J Decoursey, LBauman, M Menaker, Spectral sensitivity of a novel photoreceptive system mediating entrainment of mammalian circadian rhythms, Nature (1984) 308:186

[18] J Hannibal, P Hindersson, S M Knudsen, B Georg, J Fahrenkrug, The Photopigment Melanopsin Is Exclusively Present in Pituitary Adenylate Cyclase-Activating Polypeptide-Containing Retinal Ganglion Cells of the Retinohypothalamic Tract, J. Neuroscience (2002) 22:191

[19] J J Gooley, J Lu, T C Chou, T E Scammell, C B Saper, Melanopsin in cells of origin of the retinohypothalamic tract Nat. Neurosci. (2001) 4:1165

[20] M Maymon, T Mockler, C Lin, Blue Light-Dependent in Vivo and in Vitro Phosphorylation of Arabidopsis Cryptochrome 1, Plant Cell (2003) 15:2421

[21] C H Wu, A Nikolskaya, H Huang, L S Yeh, D A Natale, C R Vinayaka, Z Z Hu, R Mazumder, S Kumar, P Kourtesis, R S Ledley, B E Suzek, L Arminski, Y Chen, J Zhang, J L Cardenas, S Chung, J Castro-Alvear, G Dinkov, W C Barker, PIRSF: family classification system at the Protein Information Resource, Nucleic Acids Res.(2004) 1;32(Database issue):D112-4

[22] D H Haft, J D Selengut, O White, The TIGRFAMs database of protein families, Nucleic Acids Res. (2003) 31(1):371

[23] Z Melyan, E E Tarttelin, J Bellingham, R J Lucas, M W Hankins, Addition of human melanopsin renders mammalian cells photoresponsive, Nature (2005) 433:741

[24] R Hermann, L Poppe, Spilbak, C Boden, J maurer, S Weber, A Lerchl, Predicted 3D-structure of melanopsin, the non-rod, non-cone photopigment of the mammalian circadian clock, from Djungarian hamsters (Phodopus sungorus), Neuroscie Lett (2005) 376:76

[25] N J Mulder, R Apweiler, T K Attwood, A Bairoch, A Bateman, D Binns, P Bradley, P Bork, P Bucher, L Cerutti, R Copley, E Courcelle, U Das, R Durbin, W Fleischmann, J Gough, D Haft, N Harte, N Hulo, D Kahn, A Kanapin, M Krestyaninova, D Lonsdale, R Lopez, I Letunic, M Madera, J Maslen, J McDowall, A Mitchell, A N Nikolskaya, S Orchard, M Pagni, C P Ponting, E Quevillon, J Selengut, C J A Sigrist, V Silventoinen, D J Studholme, R Vaughan, C H Wu, InterPro, progress and status in 2005, InterPro, progress and status in 2005, Nucleic Acid Res (2005) 33 D201-D205

[26] J N Chen, X Zhao, J X-J Min, J X Zhang, Extracting Biologically Relevant Common Motifs from Protein Sequences, Common Motifs from Multiple Proteins (2002) 41

[27] A Gattiker, E Gasteiger, A Bairoch, ScanProsite: a reference implementation of a PROSITE scanning tool, Appl Bioinformatics (2002) 1(2):107

[28] S F Altschul, W Gish, W Miller, E W Myers, D J Lipman, Basic local alignment search tool, J. Mol. Biol. (1990) 215:40B

[29] C Bystroff , V Thorsson , D Baker, HMMSTR: a Hidden Markov Model for Local Sequence-Structure Correlations in Proteins, J. Mol. Biol. (2000) 301:173
[30] K Simons, C Kooperberg, E Huang, D Baker, Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions, J.Mol. Biol. (1997) 268:209
[31] H M Berman., J Westbrook, Z Feng, G Gilliland, T N Bhat, H Weissig, I N Shindyalov, P E Bourne, The Protein Data Bank, Nucleic Acid Res. (2000) 28(1):235
[32] I N Shindyalov, P E Bourne, Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, Protein Engineering (1998) 11:739
[33] Y Ye, A Godzik, Flexible structure alignment by chaining aligned fragment pairs allowing twists, Bioinformatics (2003) 19:ii246
[34] C Lambert, N Leonard N, X de Bolle, E Depiereux, ESYPRED3D : Prediction of proteins in 3D structures, Bioinformatics (2002) 18(9):1250
[35] J C Prasad, S R Comeau,S Vajda., C J Camacho, Consensus alignment for reliable. framework prediction in homology modeling, Bioinformatics (2003) 19:1682
[36] D E Kim, D Chivian, D Baker, Protein structure prediction and analysis using the Robetta server, Nucleic Acids Res. (2004) 32:W526
[37] G Vriend, WHAT IF: A molecular modelling and drug design program, J. Mol. Graph. (1990) 8:52
[38] G M Morris, D S Goodsell, R S Halliday, R Huey, W E Hart, R K Belew, A J Olson, Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function, J. Comput. Chemistry (1998) 19:1639
[39] C Htenyi, D van der Spoel, Efficient docking of peptides to proteins without prior knowledge of the binding site, Protein Science (2002) 11:1729
[40] D Juretic, L Zoranic, D Zucic, Basic charge clusters and predictions of membrane protein topology, J. Chem. Inf. Comput. Sci. (2002) 42:620
[41] J Schultz, R R Copley, T Doerks, C P Ponting, P Bork, SMART: A Web-based tool for the study of genetically mobile domains, Nucleic Acids Res 2000; 28: 231-234
[42] L Holm, C Sander, DALI: a network tool for protein structure comparison, Trends Biochem Sci. (1995) 20(11):478
[43] W D Ritchie, Evaluation of protein docking predictions using Hex 3.1 in CAPRI rounds 1 and 2, Proteins: Structure, Function, and Genetics (2003) 52:98
[44] D E Goll, V . Thompson, H Li, W Wei, J Cong, The Calpain System, Physiol. Rev. (2003) 83:731
[45] Y Sang, Q Li, V Rubio, Y Zhang, J Mao, X Deng, H Yang, N-Terminal Domain–Mediated Homodimerization Is Required for Photoreceptor Activity of Arabidopsis CRYPTOCHROME 1, The Plant Cell (2005) 17:1569
[46] I Friedberg, T Harder, A Godzik, JAFA: a Protein Function Annotation Meta Server, Nucleic Acids Research (2006) 34:W379

[47] P Horto., K Nakai, Better prediction of protein cellular localization sites with the k nearest neighbors classifier, Proc Int Conf Intell Syst Mol Biol. (1997) 5:147
[48] B Rost, G Yachda, J Liu, The PredictProtein Server, Nucleic Acids Research (2002) 32:W321-W326
[49] R D Finn, J Mistry, BSchuster-Böckler, S Griffiths-Jones, V Hollich, T Lassmann, S Moxon, M Marshall, A Khanna, R Durbin, S R Eddy, E L L Sonnhammer, A Bateman, Pfam: clans, web tools and services, Nucleic Acids Research (2006) Database Issue 34:D247-D251
[50] Y Harada, M Sakai, N Kurabayash., T Hirot., Y Fukada**,** Ser-557-phosphorylated mCRY2 Is Degraded upon Synergistic Phosphorylation by Glycogen Synthase Kinase-3{beta}**,** The Journal of Biochemistry (2005) 280:36, 31714-31721
[51] C Bru, E Courcelle, S Carrère, Y Beausse, S Dalmar, D Kahn, The ProDom database of protein domain families: more emphasis on 3D, Nucleic Acids Res. (2005) 33: D212-D215
[52] X Qui, T Kumbalasiri, S M Carlson, K Y Wong , V Krishna, I Provencio, D M Berson, Induction of photosensitivity by heterologous expression of melanopsin, Nature (2005) 433(7027):745-9
[53] T K Attwood, P Bradley, D R Flower, A Gaulton, N Maudling, A Mitchell, G Moulton, A Nordle, K Paine, P Taylor, A Uddin, C Zygouri, PRINTS and its automatic supplement, prePRINTS, Nucleic Acids Res. (2003) 31(1):400
[54] W J Schwartz, A clinician's primer on the circadian clock: Its localization, function, and resetting, Adv. Intern. Med.(1993) 38:81
[55] S Jones, J Thornton, Principles of Protein-Protein Interactions, Proc. Natl. Acad.Sci. USA (1996) 93:13
[56] Collection of Articles on Capri results, Proteins: Structure, Function, and Bioinformatics (2005) 60
[57] J Moult, K Fidelis, B Rost, T Hubbard, Critical assessment of methods of protein structure prediction (CASP)—Round 6, Proteins (2005) 61(S7):3
[58] D Chivian, D E Kim, L Malmstrom, P Bradley, T Robertson, P Murphy, C E Strauss, R Bonneau, C A Rohl, D Baker, Automated prediction of CASP-5 structures using the Robetta server, Proteins (2003) 53:524
[59] S F Altschul, T L Madden, A A Schaffer, J Zhang,Z Zhang, W Miller, D J Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. (1997) 25:3389
[60] L Jaroszewski, L Rychlewski, A Godzik, Improving the quality of twilight-zone alignments, Protein Sci. (2000) 9:1487
[61] L Rychlewski, L Jaroszewski, W Li, A Godzik, Strategies for structural predictions using sequence information, Science (2000) 302:1364

[62] K Ginalski, J Pas, L S Wyrwicz, M V Grotthuss, J M Bujnicki, L Rychlewski, ORFeus: detection of distant homology using sequence profiles and predicted secondary structure Nucleic Acids Res. (2003) 31:3291

[63] K Ginalski, A Elofsson, D Fischer, L Rychlewski, 3D-Jury: a simple approach to improve protein structure predictions, Bioinformatics (2003) 19:1015

[64] D.W. Ritchie, High Order Analytic Translation Matrix Elements For Real Space Six-Dimensional Polar Fourier Correlations, J. Appl. Cryst. (2005) 38:808

[65] E Katchalski-Katzir, I Shariv, M Eisenstein, A A Friesem, C Aflalo, I A Vakser, Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques, Proc. Natl. Acad. Sci. USA (1992) 89:2195

[66] S Hubbard, J Thornton, NACCESS, University College London (1993)

[67] A Gattiker, K Michoud, C Rivoire, A H Auchincloss, E Coudert, T Lima, P Kersey, M Pagni, C J A Sigrist, C Lachaize, A-L Veuthey, E Gasteiger, A Bairoch, Automatic annotation of microbial proteomes in Swiss-Prot, Comput. Biol. Chem. (2003) 27:49-58

[68] Wager-Smith K, Kay SA, Circadian rhythm genetics: from flies to mice to humans, Nature Genetics (2000) 26:23-27

[69] J Gough, K Karplus, R Hughey, C Chothia, Assignment of Homology to Genome Sequences using a Library of Hidden Markov Models that Represent all Proteins of Known Structure, J. Mol. Biol. (2001) 313(4):903

[70] C J A Sigrist, L Cerutti, N Hulo, A Gattiker, L Falquet, M Pagni, A Bairoch, P Bucher, *PROSITE*: a documented database using patterns and profiles as motif descriptors, Brief Bioinform. (2002) 3:265

[71] Turgut S, M.Sc. thesis, Koc University, 2006

[72] C L Partch, A Sancar, Photochemistry and photobiology of cryptochrome blue-light photopigments: the search for a photocycle, Photochem Photobiol. (2005) 81(6):1291

VITA

Evrim Besray Ünal was born in Istanbul, Turkey on December 13, 1981. She is an alumnus of Kültür Fen Lisesi, Istanbul. She received her Bachelor of Science degree in molecular biology and genetics from Boğaziçi University, Istanbul, in June 2004.

From 2004 to 2006, she was a research and teaching assistant in the Computational Science and Engineering Department of Koç University. Her research includes the interaction between Cryptochrome and Melanopsin proteins; Circadian rhythms.