Structure Based Analysis of the Interactions Between PER-ARNT-SIM (PAS) Domain Containing Proteins in Circadian Rhythm

by

Serap BELDAR

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

Master of Science
in
Computational Sciences and Engineering

Koc University
December 2011

Koc University

Graduate School of Sciences and Engineering

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

Serap BELDAR

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:

_____

Professor. Özlem Keskin (Advisor)

_____

Professor. Attila Gürsoy

_____

Assistant Professor Nurhan Özlü

Date: _____

ABSTRACT

Per-ARNT- Sim (PAS) domains are modular protein units those are critical for regulation of clock-controlled gene expression. The mammalian BMAL1 and CLOCK are the transcription factors that contain two basic helix-loop-helix domains and bind E-box elements (CACGTG) of clock regulated genes including the Period and Cryptochrome and activate their transcription. Then the PERIOD (PER) and CRYPTOCHROME (CRY) proteins form ternary complexes with casein kinase Iε (CKIε) in the cytoplasm and translocate into the nucleus, where they act as a negative regulator of BMAL1/CLOCK-driven transcription. To understand the nature of interaction between PER2-CLOCK-BMAL1 complex, we performed structure based analysis on protein-protein interactions (PPI) those formed *via* PAS domains of clock proteins. This complex was analyzed by using structural data and efficient structural comparison algorithms to predict potential interactions. Since there are no available atomic structures for our proteins of interest, homology models are used. In our model, BMAL1 and CLOCK interacts with each other through their PAS B domains and the PER2 interacts with this dimer through the PAS B domain of the CLOCK. On BMAL1/CLOCK interface, we found 12 hotspots, 7 residues on BMAL1 (347,349,362,405,427,429,441), 5 residues on CLOCK (317, 338, 350, 352,376). On PER2/CLOCK interface, 8 hotspots are found, 3 residues on PER2 (414, 429,431) and 5 residues on CLOCK (332,354,356,333,361). Here we show how, using structural data and efficient comparison algorithms can explain forming the clock complexes at the molecular level. This study is not only important to understand clock mechanism at structural level but also will allow us to develop drugs against clock-regulated diseases, like Jet-Lag and some forms of depression.

ÖZET

Per-ARNT-Sim domainleri zaman kontrollü genlerin düzenlenmesinde çok önemli rolleri olan modüler protein birimleridir. Memeli transkripsiyon faktörlerinden BMAL1 ve CLOCK iki sarmal domainine sahiptir ve Period ve Cryptochrome zaman genlerinde bulunan E-box elementine (CACGTG) bağlanarak onların ekspresyonunu arttırır (geri besleme döngüsünün pozitif kolu). PERIOD (PER) and CRYPTOCHROME (CRY) proteinleri ise the BMAL1/CLOCK heterodimerinin arttırdığı ekspresyonu engeller (geri besleme döngüsünün negatif kolu). PER ve CRY sitoplazmada kasein kinaz Iε (CKIε) ile birlikte üçlü bir kompleks oluşturur. Bu üçlü yapı hücre çekirdeğine yerleşerek BMAL1/CLOCK aktifleştirdiği genleri susturur. Bu üçlü kompleksin yapısını anlayabilmek için proteinlerinin yapısında bulunan PAS domainleri arasındaki protein - protein etkileşimi üzerine yapısal analizler yaptık. Bu çoklu yapı, yapısal data ve potansiyel etkileşimi tahmin edebilen etkili karşılaştırma algoritmaları kullanılarak analiz edildi. Elimizde ilgilendiğimiz proteinlerin atomik yapıları bulunmadığından bu proteinlerin homolog modelleri yapısal data olarak kullanıldı. Modelimize göre BMAL1 ve CLOCK proteinleri birbirleriyle PAS B domainleriyle etkileşime giriyor ve PER2 bu ikili yapıya CLOCK'un PAS domaini aracılığıyla bağlanıyor. BMAL1/CLOCK etkileşim yüzeyinde 12 önemli amino asit bulduk. 7 tanesi (347,349,362,405,427,429,441) BMAL1'de ve 5 tanesi (317, 338, 350, 352,376) CLOCK üzerinde. PER2/CLOCK yüzeyinde 8 önemli aminoasit bulduk, 3 tanesi (414, 429,431) PER2'de ve 5 tanesi (332,354,356,333,361) CLOCK üzerinde. Bu çalışmada yapısal verinin ve verimli karşılaştırma algoritmalarının, biyolojik saat proteinlerinin moleküler boyuttaki etkileşimlerini nasıl açıklayacağını gösterdik. Bu çalışma sadece zaman mekanizmasınının etkileşimlerini moleküler boyutta göstermekle kalmayıp zaman proteinlerinin düzenlemesinin bozulmasıyla oluşan Jet-Lag ve bazı depresyon çeşitlerine karşı ilaç geliştirilmesinde yardımcı olacaktır.

ACKNOWLEDGEMENTS

I would like to thank my advisors **Prof. Özlem Keskin**, **Prof. Attila Gürsoy** and **Assoc. Prof. I. Halil Kavaklı** for providing me this opportunity for graduate research. It was a great chance for me to study with them- their experience and support provided me a valuable work experience in interdisciplinary era.

I am particularly grateful to **Gözde Kar, Engin Çukuroğlu** and **Özge Engin** for their generous help regarding to the computational procedures and motivation throughout my graduate study.

Thanks to all people in my office, **Emine Güven, Ömer An, Büşra Topal, Tuğçe Yıldızoğlu, Halil Peynirci** and **Pelin Atıcı** for their friendly integration and support, especially. I will never forget all the cheerful moments, we spent together.

Finally, I thank to **my family** who gave me their never- ending support and patience during every moment of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

bHLH	basic helip-loop-helix
BMAL1	Brain-muscle-Arndt-like-protein-1
*CK1*ε	Casein kinase 1 epsilon
CLOCK	Circadian-locomoter-output-cycles-protein
CRY	Cryptochrome
CYC	Cycle
DNA	deoxyribonucleic acid
PAS	Per-Arnt-Sim
PDB	Protein Data Bank
PER	Period
PPI	Protein-Protein Interaction
PRISM	Protein Interactions by Structural Matching
*REV-ERB*α	V-erbα-related protein
Ror-α	Retinoic acid receptor alpha
SCN	suprachiasmatic nucleus

Chapter 1

1 INTRODUCTION

Circadian rhythms are the intrinsic time keepers that organisms use to measure time and anticipate changes in environment such as light and temperature. Anticipatory behavior provided by circadian clock is very essential for an organism to increase the efficiency of physiological processes. When an endogenous time keeper adapts to environmental alterations, fitness and adaptation are significantly increased [1, 2]. That is why the clock has been found to exist in all prokaryotic and eukaryotic organisms, from cyanobacteria to vertebrates.

In mice and humans this rhythmic behavior is generated by a molecular clock with a periodicity of about 24 h that consists of a transcription-translation feedback loop (TTFL). The four clock proteins are CLOCK, BMAL1, PER and CRY that are responsible for the maintaining and generation of the clock. CLOCK and BMAL1 have a basic helix loop helix (bHLH) - Per-ARNT- Sim (PAS) domain, are hetrodimerized in cytoplasm and translocate into nucleus where they mediate transcriptional activation of target genes that present E-box *cis*-regulatory enhancer on their promoter which form the positive elements of the core loop [3-5]. These positively regulated genes include *Periods, Crys, Rev-Erb, Ror-a* and several other clock-controlled genes. In the cytoplasm, PER proteins form complexes with CRYs, which then translocate in to the nucleus. Once the complexes reach their destination, the trimeric complexes inhibit the CLOCK-BMAL1 driven transcription, thereby establishing a negative feedback loop [6, 7]. The suppressed E-box driven transcription leads to a decline in PER and CRY protein levels which leads to a re-activation of CLOCK-BMAL1 driven transcription and the next circadian cycle starts.

CLOCK, BMAL1 and PER contain PAS (Per-ARNT- Sim) domains, except CRY. PAS domains are modular protein units those are the critical elements of the signal transduction and protein-protein interaction (PPI) processes. Many proteins contain multiple PAS domains that are tandemly repeated, but the role of multiple PAS domains is poorly understood, such as, how do multiple PAS domains form a heterodimer? Understanding of the structural basis of multiple PAS domain heterodimerization is very essential because it affects not only the DNA binding characteristic of dimmers but also interaction with other proteins that resulted in alteration of gene transcription regulation.

Many different approaches have been employed to characterize the interactions between the repressors PER2, CRY1 and CRY2 and the BMAL1-CLOCK heterodimer, but still the exact model is far from being apparent at the structural level. It is known that CRY1/2 plays a key role in repressing the transcriptional activation potential of the BMAL1-CLOCK heterodimer [8-11].

This study is focused on studying the structural mechanism regarding to how the core circadian clock elements of mice form a complex, including (i) how protein-protein interactions among three PAS domain containing core clock proteins form, CLOCK-BMAL1 heterodimerization and CLOCK-BMAL1-PER2 trimeric complex formation (ii) elucidating the essential residues for these PPIs. Further we also model the interaction between CRY and BMAL1-CLOCK-PER2 complex.

In the next part, chapter 2, the literature regarding to circadian rhythm focusing on the molecular interactions within clock proteins is presented. Background of the circadian rhythm, the detailed information of each core circadian clock components, PAS domains and molecular interactions between clock proteins are explained.

Chapter 3 includes explanation of datasets and the methodology that is used to construct this thesis. Here, first the details about dataset are given. Then the methodology is explained and it is illustrated with a workflow.

In chapter 4, the model structures of complexes are presented. First, the BMAL1-CLOCK-PER2 complex which is predicted by PRISM algorithm is illustrated. Then, The CRY-BMAL1-CLOCK-PER2 complex model is demonstrated. In the final part of this chapter, the critical protein interfaces and hotspots on interface residues are shown.

The last chapter ends with discussing the results, presented in chapter 4.

Chapter 2

2 LITERATURE REVIEW

This chapter comprises the summary of the detailed review of the literature about general features of circadian rhythm. In the first part, general molecular regulation of the circadian clock machinery is explained. Then, control of circadian clock on molecular level and molecular interactions within clock components is reviewed. In the last part, common properties of PAS (Per- ARNT- Sim) domains are described.

2.1 Overview Circadian Rhythm

Any living organism on earth is subjected to rhythmic changes of its environment and the major rhythm influencing the earth is rotation of earth around the sun. Environmental changes make the biological organisms to adapt to the daily alterations between light and dark, and high and low temperatures. Those adaptations led to the evolution of an endogenous time keeper, known as the circadian clock. This biological timing system helps the organism to synchronize developmental and metabolic events to the most suitable time of the day. For example, in human, endogenous clock regulates the sleep/wake time, body temperature and endocrine cycles of human beings. It is believed that circadian rhythms evolved due to these predictable rhythms and have been fine-tuned under selective pressure [12].

Circadian rhythm is the cycle of 24 hour biological processes that allow appropriate timing of physiology and behavior of biological systems, by that way it optimizes the efficiency of metabolism. Circadian rhythms are composed of 3 main elements. First the oscillations are initiated by environmental cues, called as inputs. The input signal can be light, food intake, chemical or physical factors. The input signals are received via receptors.

Second the clock molecular components generate biological rhythm in central oscillator and lastly the clock information is transmitted to other pathways that provide the periodicity of the rest of the organism[13].

Any environmental input that has the ability to reset the clock is called a Zeitgeber (time giver). The most obvious Zeitgebers is light [14]. One other important feature of the circadian rhythm is temperature reparation. It is known that the rates of biochemical reactions are severely affected by the varying temperature. However, interestingly, circadian machinery is unaffected by varying temperature, known as temperature compensation, though, daily temperature changes can also act as Zeitgebers in plants [15]. This feature common to all circadian system provide a basis for stability in multicellular adaptation in organisms. Even though this feature implies independence of circadian systems from external factors, circadian clock is controlled by environmental factors such as change in light intensity, nutrient availability and social pressure to be able to provide adaptive advantage [16]. This is also a common feature of all circadian systems and generally called “resetting of the circadian clock”. For a circadian clock machinery to be useful in environmental adaptation, it should be entrained or reset by the periodic stimuli from various environmental factors. The process by which the biological clock re-sets itself following particular environmental signals is called entrainment.

2.2 Mammalian Circadian Clock

The mammalian circadian system is organized in a hierarchy of oscillators. Suprachiasmatic nucleus (SCN) of the anterior hypothalamus is present at the top of this hierarchy. The SCN is responsible for coordinating independent peripheral oscillators so that a coherent rhythm is orchestrated at the organismal level. In **Figure 2-1** the physiological organization mammalian clock is shown [17-19].

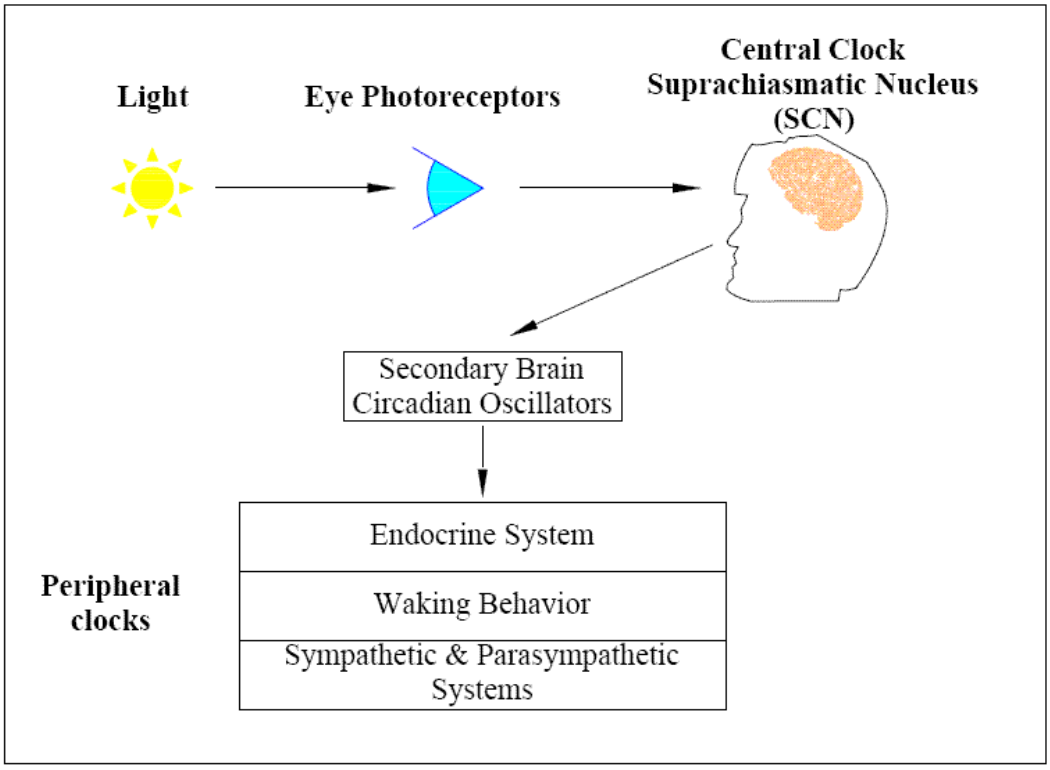


Figure 2-1 Hierarchical organization of the mammalian circadian clock. The suprachiasmatic nucleus (SCN), located in the anterior hypothalamus, receives light stimuli from the retina (eye receptors) via the retinohypothalamic tract. The oscillation in SCN neurons is entrained by light stimulus and results in a humoral or neural output that is received and interpreted by numerous peripheral clocks such as endocrine system. Peripheral clocks produce periodic transcriptional and posttranslational responses that ultimately produce behavioral, metabolic, and physiological output.

The clock mechanism in the SCN and the peripheral oscillators are known to be similar at the molecular level, which consists of a network of transcriptional–translational feedback loops that drive rhythmic, ~24-h expression patterns of core clock components. Core clock components are defined as genes whose protein products are necessary for the

generation and regulation of circadian rhythms within individual cells throughout the organism. There are four major components that mediate the core clock, CLOCK; BMAL1, PER and CRY proteins. The two positive components of the feedback loop are CLOCK and a heterodimeric binding partner, BMAL1. Each of them contains protein sequences for both basic helix loop helix (bHLH) and PAS (PER-ARNT-SIM) dimerization domains. According to this model, the proteins encoded by these Clock and Bmal1 genes dimerize through the bHLH and/ or PAS domains, bind regulatory DNA sequences (E-boxes) and activate transcription of three *Period* genes (mPer1, mPer2, and mPer3) and 2 *Cryptochrome* genes (mCry1 and mCry2) and other clock regulated genes [6, 7]. Once PER2 and CRY proteins are translated, they reside at cytoplasm for a while, the exact staying time in cytoplasm is determined by the concentration of PER and CRY proteins. When the concentration of PER2 is reached a particular amount, CRY and casein kinase l epsilon (CKIε) proteins form a multimeric complex and PER2 is phosphorylated by casein kinase l epsilon (CKIε).

It is shown that CRY protein stabilizes the interaction between PER2 and casein kinase l epsilon [20-22]. (shown in **Figure 2-2**) Then phosphorylated PER2 proteins as in a complex with CRY are accumulated in nucleus and subsequently they interact with CLOCK-BMAL1 complex to inhibit their own transcription. In time, the PERIOD and CRY proteins turn over, their negative feedback is reduced, and the CLOCK-BMAL1 heterodimer can begin transcription again. This represents one circadian cycle of the clock, as the complete time course of this feedback loop is ~24 hours.

2.3 Molecular Regulation of the Circadian Clock

The presence of an internal timekeeper is a universal biological process conserved throughout evolution. Biological rhythms have been observed in most species including

unicellular organisms, plants, insects, birds, and mammals. As a result, the behavioral and physiological rhythms controlled by the clock are as diverse as the organisms in which they are found. In humans, for instance, rhythms in sleep-wake cycles, body temperature, hormonal secretions, blood pressure, and liver metabolism are just a few examples of those biological processes regulated by circadian clock. Although biological rhythms oscillate with various frequencies, in various processes; it has highly conserved molecular components in all organisms.

A major area of interest related to biological clock is to understand the regulatory mechanisms of the clock proteins at molecular level. Forward and reverse genetic approaches have successfully identified the molecular components of biological clock [15]. The first clock gene, *per*, was identified in *Drosophila* [23]. After that, several other clock genes were identified in various eukaryotic organisms.

The transcriptional and translational feedback loop (TTFL) model was proposed to explain the oscillation of mRNA and proteins of clock genes [24]. In this model, promoters of clock genes are under the control of positive transcriptional factors, whose function is inhibited by post-translationally modified clock proteins. In other words, the protein products of the clock genes negatively feedbacks on their own expression in an auto-regulatory feedback loop. **Figure 2-2** shows the schematic of a TTFL model.

This negative feedback loop was considered as a chemical equilibrium, in which the rhythmic concentrations of the negative transcriptional elements induce transcriptional repression which implies the exact repression time of positive transcription factors is determined by the concentration of negative elements. The temporal increase in negative elements' concentration controls the negative feedback interactions. However it is shown that not only the temporal increase in negative factors' concentration is the limiting step,

but also post transcriptional regulations on negative elements determine the timing of circadian clock. Phoshorylation is one of the very critical post translational modifications for regulation of circadian rhythm. Kinases CKIε, CKIδ and GSK3β have been shown to involve in regulation of both the positive and the negative transcriptional elements [20, 21, 25]. In the next part, the core clock proteins will be reviewed regarding to their structures and functionality.

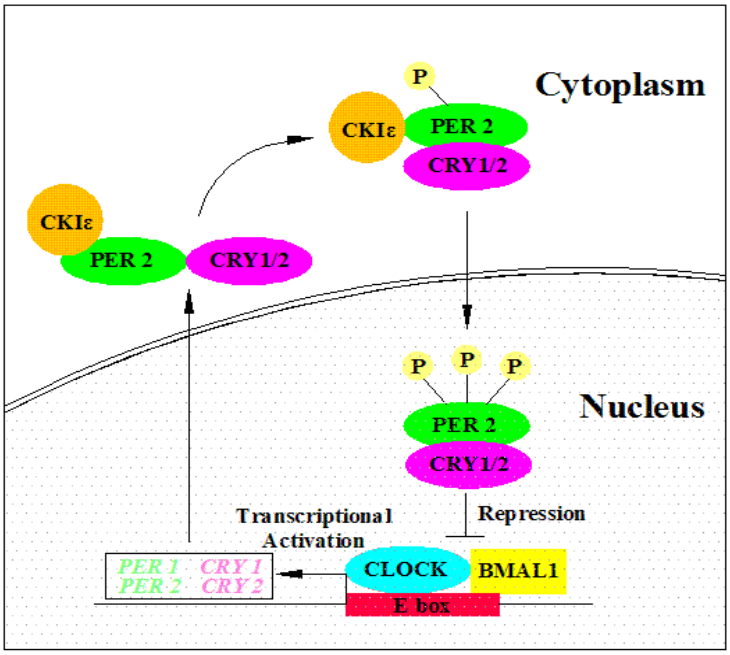


Figure 2-2 Molecular regulation of mammalian circadian clock.

2.3.1 CLOCK

Among all of the mammalian circadian rhythm–related genes, *Clock* (*Circadian Locomotor Output Cycles Kaput*) is unique by having been first characterized by a forward

genetics strategy [26]. *Clock* gene encodes the basic-helix-loop-helix domain (bHLH) and PAS domain (PER-ARNT-SIM) family of transcription factors [27, 28]. The mouse CLOCK protein is 855 amino acids in length and in addition to these DNA binding and protein interaction domains, it also includes a glutamine-rich C-terminal domain, demonstrated in other transcription factors to be important for transactivation capabilities **Figure 2-3** shows the cartoon representation of domains of CLOCK protein.

Analysis of Clock knockout mice indicated that, disruption of *Clock* does not change the circadian oscillations. This implies that *Clock* gene product is redundant and not essential for generation of circadian rhythms in mammals [29]. The inconsistency between these two studies on *Clock* gene disruption possibly arises from a dominant negative effect of CLOCKΔ19 which competes in binding to other protein factors. It has been suggested that this redundancy occurs since NPAS2 protein also mediates similar functions as CLOCK [30]. CLOCK also has a chromatin remodeling capability, which results from the histone acetyltransferase activity that is mediated through C-terminus of CLOCK [31]. Histone remodeling activity of the CLOCK has been shown to be an essential component for transcriptional activity, which was long thought to have a role in the circadian clock mechanism.

Another important feature of the CLOCK protein is that it is not transcribed or translated in a rhythmic way during the circadian period. Clock proteins are always present during circadian regulation [20]. It means that, the rhythmic expression of BMAL1 is the main determinant in formation of the rhythms for the positive part of the circadian clock.

2.3.2 BMAL1

BMAL1 gene is also known as MOP3 and ARNTL (Aryl hydrocarbon receptor nuclear translocator-like). The mouse BMAL1 protein is 632 amino acids long and it only

contains DNA binding and PAS domains [32]. (See **Figure 2-3**) The protein transcribed by this gene has a basic-helix-loop-helix (DNA binding) and PAS (PER- ARNTL-SIM), domain that forms a heterodimer with a second bHLH-PAS protein, CLOCK, or its ortholog, Npas2 to activate transcription of PER and CRY [3, 33].

In addition to mammals like mice and human, homologs of the *Arntl* gene are found in lower organisms like fish, bird and Drosophila (the fly homolog is known as *Cycle* gene). BMAL1 is the component of the mammalian circadian rhythm whose deficiency in mice model resulted in arrhythmicity [34]. It is shown that loss of the PAS protein of BMAL1 in mice results in immediate and complete loss of circadian rhythmicity in constant darkness. Additionally, locomotor activity in light-dark (LD) cycles is impaired and activity levels are reduced in *Bmal1* -/- mice. Moreover *Bmal1* null mouse has a variety of other physiological abnormalities, indicating that BMAL1 is not only an essential component of the circadian clock, but also an important mediator of physiological events [35].

In contrast to its interacting partner CLOCK which has a constitutive promoter and always present in the nucleus, *Bmal1* gene regulation oscillates like circadian cycle. *Bmal1* transcription is reciprocally regulated by the orphan nuclear receptors NR1D1 (Rev-erb-α) and NR1F1 (ROR-α) which form a second interlocking loop in the mammalian circadian clock [36-39].The other nuclear receptors of the same families (NR1D2 or Rev-erb-β; NR1F2 or ROR-β; NR1F3 or ROR-γ) were also shown to act on *Arntl* gene [39].

2.3.3 PERIOD

Genes in this family encode components of the circadian rhythms of locomotors activity, metabolism, and behavior and consists of three genes (PER1, PER2, and PER3) This gene family is expressed in a circadian pattern in the suprachiasmatic nucleus, the

center circadian pacemaker in the mammalian brain. Circadian expression of *Period* in the suprachiasmatic nucleus continues in constant darkness, and a shift in the light/dark cycle evokes a proportional shift of gene expression in the suprachiasmatic nucleus.

Period gene was first identified in Drosophila by using forward genetics screen approach [40]. The mouse PERIOD protein is 1253 amino acids long and it has a variety of structural domains, basic HLH domain in the N-terminus, followed by two repetitive PAS domains and an additional PAS-like domain. (**Figure 2-3**) PERIOD orthologs contain two PAS domains which are also present in BMAL1 and CLOCK. Protein-protein interactions are known to be mediated through the PAS domains between the core circadian clock proteins. These two PAS domains, PAS A and PAS B, are present as tandem repeats in the N terminus of PERIOD. Elimination of one of these PAS domains (PAS B) in m*Per2* results in arrhythmicity in constant darkness in mice[41]. In the C terminal of PERIOD, there is a coiled coil region which has functional significance since this region is found be as the binding site for CRY-PERIOD interaction region [42].

During the regulation of clock by transcriptional feedback loop of the core circadian clock system, PERIOD associates with CRY proteins which then interact to form a protein complex to create the negative limb of the feedback loop [1]. Knockout studies showed that, mice with a null allele of m*Per1* or m*Per2* has a shorter circadian rhythm, and m*Per2* knockout mice resulted in a complete disruption of circadian rhythmicity [43]. On the contrary, mice knockout mice ,m*Per3 -/-* display a subtle shortened circadian period with nearly normal circadian function [44]. These findings imply that PERIOD paralogs have different roles in clock function and m*Per2* is the essential component in the negative limb of core feedback loop.

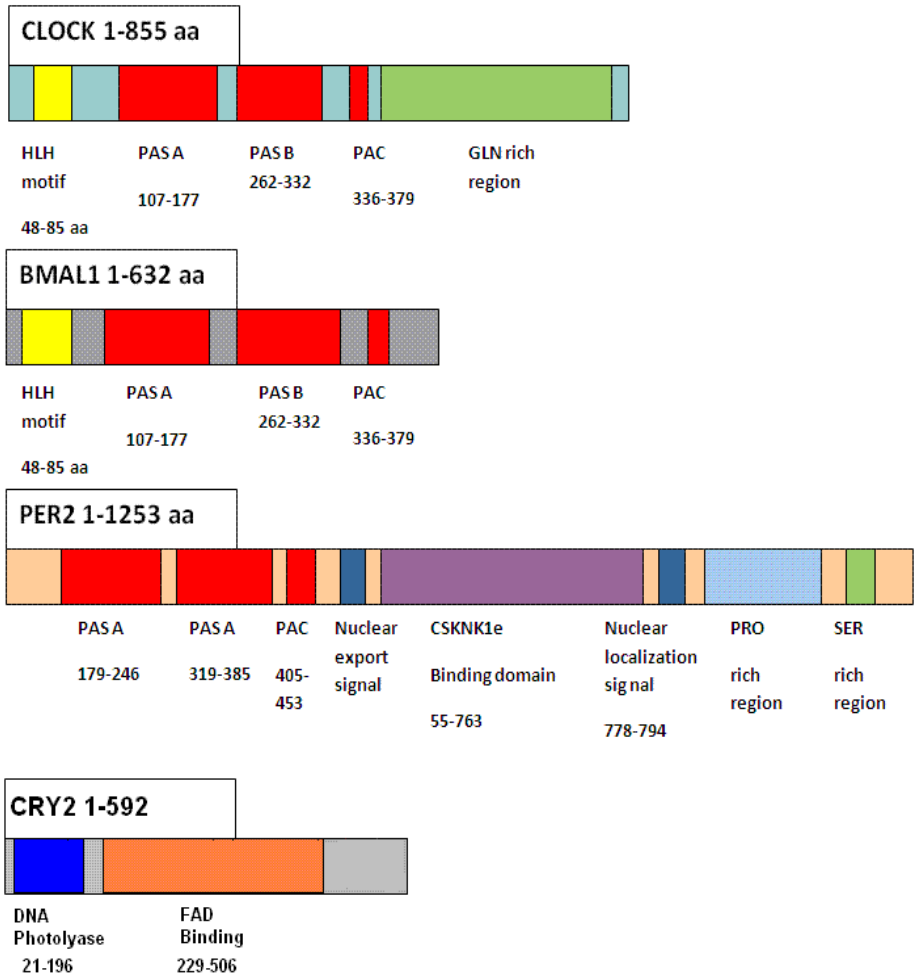


Figure 2-3 Proteins of the mammalian circadian clock. Schematic diagrams of the full-length proteins identified in mammals that contain PAS domains .The proteins are retrieved from UniProt Database (mouse sequences are selected). CLOCK, Clock (circadian locomotor output cycles kaput, UniProt Accession O08785 (CLOCK_MOUSE); BMAL1, (ARNT-like).Uniprot Accession Q9WTL8 (BMAL1_MOUSE); PER2 (Period2), Uniprot Accession O54943 (PER2_MOUSE).

2.4 PAS Domains

PAS domains are modular protein units that are the critical elements of the signal transduction process in the regulation of clock-controlled gene expression. The Per-Arnt-Sim(PAS) was originally defined as ~275 amino acids of sequences in *Drosophila melanogaster*- clock protein *period(per)*, in flies- a neuro developmental regulator *single-minded* (*sim*); and in human- a component of the dioxin signaling pathway aryl hydrocarbon receptor nuclear translocator(ARNTL)[23]. By explosion of genome sequence information, new thousands of PAS domain containing proteins were identified in other species, from Archaea, bacteria and Eukaryote [45].

The definition of PAS domain has changed with the characterization of more PAS containing domains. It is seen that there are ~70 amino acids of repeating sequences called as PAS A and PAS B also known as PAS1 and PAS2. PAS domains are often associated with PAC domains which are the C terminal of PAS domains. It appears that these domains are directly linked; together they form the conserved 3D PAS fold.[46] The division between the PAS and PAC domains is caused by major differences in sequences in the region connecting these two motifs. According to the phylogenetic analysis of PAS domains, there is an obvious sequence difference between PAS A and PAS B domains. The sequence similarity within PAS A domains is 35% and it is 31% in PAS B domains. Whereas, a PAS A and a PAS B domain can be identical to each other at most 20% [47]. In contrast to functional diversity, the structure of the core of PAS domains is mostly conserved. The conserved structure of a PAS fold comprises a central antiparallel β sheet with five strands and several numbers of α helices flanking the sheet [48]. Signals originate within the conserved structure of PAS domains, and then they generate structural and dynamic changes predominantly within the β sheet. In spite of low sequence similarity of PAS domain, a high conservation in the structure fold and topology of known structures

suggest a strong evolutionary conservation of some functional features. Since PAS domains are very critical in signal transduction, high conservation in structures provides the stable protein interfaces.

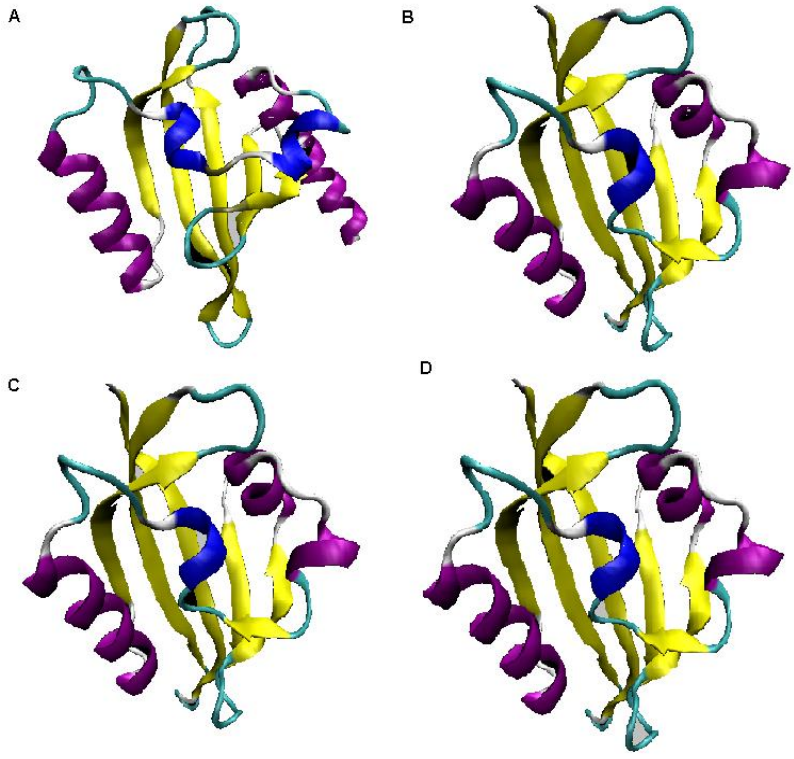


Figure 2-4 Conservation folds within PAS domains.Cartoon representation of the PAS secondary structures are shown here. It is seen that, a canonical PAS fold comprises a central antiparallel β sheet with five strands and several numbers of α helices flanking the sheet. (A) NifL (PDB i.d.: 2GJ3 chain A),(B) HERG (PDB i.d.: 1BYW), (C) FixL (PDB i.d.: 1DRM), (D)hPASK (PDB i.d.: 1LL8) .

PAS domains are involved in many signaling pathways where they are used as a signal sensor domain. For that reason PAS domain have functions in adaptive pathways like hypoxia response pathway, polycyclic aromatic carbons and the dioxin pathway and mammalian circadian clock [49, 50]. PAS sensors detect chemical and physical stimuli and regulate the activity of functionally diverse effectors domains like circadian clock or hypoxia pathway.

PAS structures derived by NMR spectroscopy improved the structural knowledge of PAS fold. In **Figure 2-4,** secondary structures of PAS domain of the protein NifL from Azotobacter vinelandii (2GJ3), N-terminal helices of HERG potassium channel, a eukaryotic PAS domain. (1BYW), FixL protein in B. japonicum (1DRM), and N terminal of PAS kinase from H. Sapiens (2J80) are demonstrated to provide more clear understanding on PAS domains conserved fold.

Many PAS domains involved in biological signaling pathways have been identified, however only limited structural information is available regarding how they interact with each other in multiprotein complexes. One of the most studied process is the hypoxia response pathway, which allows eukaryotic cells to respond to low oxygen tension via the formation of a heterodimeric complex between ARNT and another bHLH-PAS protein, the hypoxia-inducible factor alpha (HIF-a). It has been shown that both PAS domains of HIF-2a are required for HIF activity in living cells, and also that both contribute to the heterodimerization. In addition to this, it is shown that point mutations on the HIF-2a PAS-B domain disrupt the interaction between HIF-2a and ARNT PAS-B which indicates that PAS-B domain of HIF-2a is the key region for the interaction between HIF-2a and ARNT [51] (see **Figure 2-5**). While several crystal structures of PAS domain homodimer have been solved to date [52-54] , structures of PAS domain of heterodimeric complexes have been particularly less resolved. One of the most comprehensive study that analyze the

heterodimeric complex of PAS domain is done by HADDOCK and site-directed spin labeling approaches[55]. This work shows that the HIF-2a and ARNT PAS-B domains interact via an antiparallel association of the solvent-exposed surface of their central β-sheets.

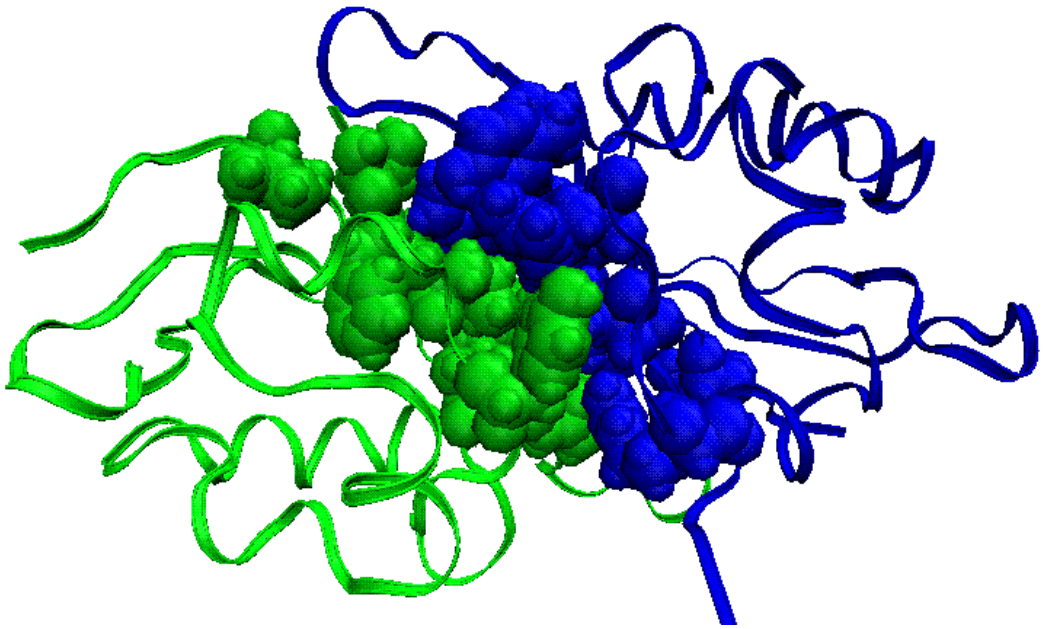


Figure 2-5 Structural basis for PAS-PAS heterodimerization. Ribbon diagram of the HADDOCK-derived structure of HIF-2α/ARNT PAS-B domain complex. Ball shapes show the interfaces and it seen that they are located on β- sheet region.

2.5 Protein Interactions between BMAL1-CLOCK-PER2-CRY2

Many different approaches have been employed to characterize the interactions between the repressors PER2, CRY1 and CRY2 and the BMAL1-CLOCK heterodimer to

understand the repression mechanism. In a recent study[56], physical interactions between repressors (PER2 ,CRY) and CLOCK-BMAL1heterodimer with E-box (DNA on) was analyzed. It is shown that CLOCK-BMAL1 binds to an E-box sequence in DNA and CRY binds to the CLOCK: BMAL1: E-box ternary complex separately from PER2, where as PER2 does not. It is found that CRY plays a key role in repressing the transcriptional activation potential of the heterodimer, and recently, various attempts have been made to elucidate the mechanism by which interaction of CRY with BMAL1 and/or CLOCK inhibits transcription [8-11]. However, many of these studies use multimeric protein complexes, which do not always satisfactorily identify the exact interaction regions between two individual components of the complex. In next section, the protein interaction studies between clock components will be reviewed.

BMAL1 was found as binding partner of CLOCK by using protein interaction and transactivational assays. DNA binding domain (bHLH) of BMAL1-CLOCK heterodimer bind six base-pair DNA sequence (CACGTG), which is also known as E-box [57]. It was shown that co-expression of CLOCK and BMAL1 lead to protein binding of a *Drosophila period* gene enhancer that is required for circadian rhythm [58, 59]. Analysis of the mPer1 promoter revealed the presence of three E-box sequences in the 1.2 kb upstream of the mPer1 transcription start site. When the E-box sequences were randomly disrupted, no transcriptional activity was detected. These experiments also demonstrated that CLOCK and BMAL1 stimulates the transcription of mPer1.

In another study, the importance of PAS domains within BMAL1 and CLOCK for heterodimerization was shown Only separate PAS domain regions (PAS A, PAS B and PAS AB) from BMAL1 and CLOCK were cloned and binary interactions within these 6 clones were tested. It is been shown that a single PAS domains cannot sustain heterodimerization alone. Both PAS AB domains from BMAL1 and CLOCK are required

for heterodimerization. In addition to this, it is seen that PAS B domain of BMAL1 can contribute interaction with CLOCK PAS AB [9].

In several studies, it is found that CRY2 protein binds to BMAL1 through PAS B domain of BMAL1 [9, 10, 60, 61] . The interaction between CRY and BMAL1 is thought to be very critical for negative control of circadian rhythm. The importance of mBMAL1 for clock function is clearly demonstrated by the fact, that mBMAL1-/- knockout mice show an immediate and complete loss of circadian rhythmicity [34]. The C-terminal coiled coil and tail (CCtail) regions of the mammalian cryptochromes (mCRY1/2) and the C-terminal mBMAL1 region critically regulate the activity of the mBMAL1/mCLOCK transcription factor complex within the mammalian circadian clock [8]. It is known that the residues from 369 to 488 on CRY2 are required for inhibition of CLOCK-BMAL mediated transcription.

The interaction between BMAL1 and PER2 was observed by using mammalian two hybrid system [21, 62] . Additionally, it was seen that when PAS A domain of BMAL1 is deleted, interaction intensity between BMAL1 and PER2 decreases which implies PER2 binds the PAS A domain of BMAL1 [60]. However when CRY proteins are added this complex, CRY1 and CRY2 influence the interaction between PER2 and BMAL1. The CRY proteins moreover seem to have a higher affinity to BMAL1 than PER2. When PER2 is added to BMAL-CRY complex, it is seen that, PER2 does not significantly affect the interaction between CRY and BMAL1 [9].

The binary interaction between CLOCK and PER2 is shown in many studies [21, 60, 62, 63]. However, the interaction regions within those proteins those are required for binding are not studied neither at domain level nor residual level.

Interaction between CRY and CLOCK proteins cannot be found [60]. However, contradictory findings have been published. Griffin et al. do not observe clear interactions between CRY1/2 and CLOCK in a yeast two-hybrid system [6] whereas Shearman et al. do [22]. Kiyohara et al. [10] report that they were not able to co-immunoprecipitate CLOCK with CRY1 in the absence of BMAL1, and also in other cases BMAL1/CLOCK(/PER2)/CRY1 complexes .Therefore, it appears likely that in this case the respective results depend strongly on the system and cell type used and that cell-specific factors might be involved in mediating the interaction. In the mammalian two-hybrid system, Langmesser et al. performed in HER911 cells and they found that CRY1 and 2 do not interact with CLOCK [60].

The regulation of circadian clock in mammals is based on a transcriptional-translational feedback loop. In this regulation mechanism BMAL1- CLOCK heterodimer directly bind E-box of DNA and then PER2-CRY complex inhibit the interactions within BMAL1- CLOCK + E-box complex and repress their own transcription. In a recent study, the physical interactions between BMAL1- CLOCK + E-box complex and repressor protein PER2 and CRY are analyzed by using both in vivo and in vitro methods [56].

They found that CRY binds stably to the CLOCK-BMAL1+E-box ternary complex independently of PER. Both CRY and PER bind to CLOCK and BMAL1 off DNA but, contrary to CRY, PER does not bind to the CLOCK-BMAL1-E-box complex. In addition to this, it is observed that PER decreases the binding affinity of CRY to the CLOCK:BMAL1:E-box ternary complex.

Chapter 3

3 Structure Based Modeling of the Clock Proteins

The current model for the circadian clock in mammals is explained by a transcription- translation feedback loop, which is reviewed in Chapter 2. In this model, CRY and PER repress their own transcription by suppressing the transactivation function of CLOCK-BMAL1 heterodimer directly, which implies there are some physical interactions. In this chapter, the methodology which is applied to model the physical interactions within clock components is presented. Here, we follow a structure based analysis of protein interactions within clock proteins. We aim to analyze the interactions by using structural data and efficient structural comparison algorithms. Since, there are no available atomic structures for our proteins of interest, we used homology models as structural data and they are used in high performance PPI algorithm as targets to model the formation of this multimeric complex. This work comprises 5 main computational analysis steps which are homology modeling, clock pathway specific template construction from PDB, high performance prediction algorithm (PRISM)[64], interface-hotspot analysis (Hotpoint)[65], KFC2 (Knowledge-based FADE and Contacts) server[66] and mutation analysis (FoldX)[67]. Our methodology is simple but very promising to understand clock mechanism at molecular level with its structure based analysis approach. In the next part, the dataset used in this work will be presented.

3.1 Dataset

3.1.1 Protein Sequences

In this work, mouse circadian rhythm is chosen for analysis. For that reason, mouse clock protein sequences are used in this methodology. Protein sequences are retrieved from Reviewed,UniProtKB/Swiss-Protdatabase [68]. The sequences and their UniProt accession

numbers are BMAL1-Q9WTL8(BMAL1_MOUSE), CLOCK-O08785(CLOCK_MOUSE), PER2O54943(PER2_MOUSE) and CRY2-Q9R194.

3.1.2 Template Set for Protein Interaction by Structural Matching (PRISM) Algorithm

An interface is the contact region between two interacting proteins. The template set represents the subset of structurally non-redundant interface architectures. Usage of the known interfaces provides the prediction of new interacting protein pairs. For this work, we used two different template sets. Default template set(1037-heterodimer set) and case-specific created template set from PDB [69]. **Table 3-1** shows the clock pathway specific created template set and in Appendix A the structures of the template set those belong to different type of interfaces are visualized.

3.2 Methodology for Structure Based Modeling of Clock Proteins

In this work, the mechanisms that underlie the formation of protein -protein interaction through their PAS domains within clock proteins is major interest of this work. Particularly CLOCK, BMAL1 and PER proteins contain PAS domains, therefore we focus on the BMAL1- CLOCK heterodimer formation and CLOCK, BMAL1, PER multimeric complex formation. Here we analyze the interactions by using structural data and efficient structural comparison algorithms.

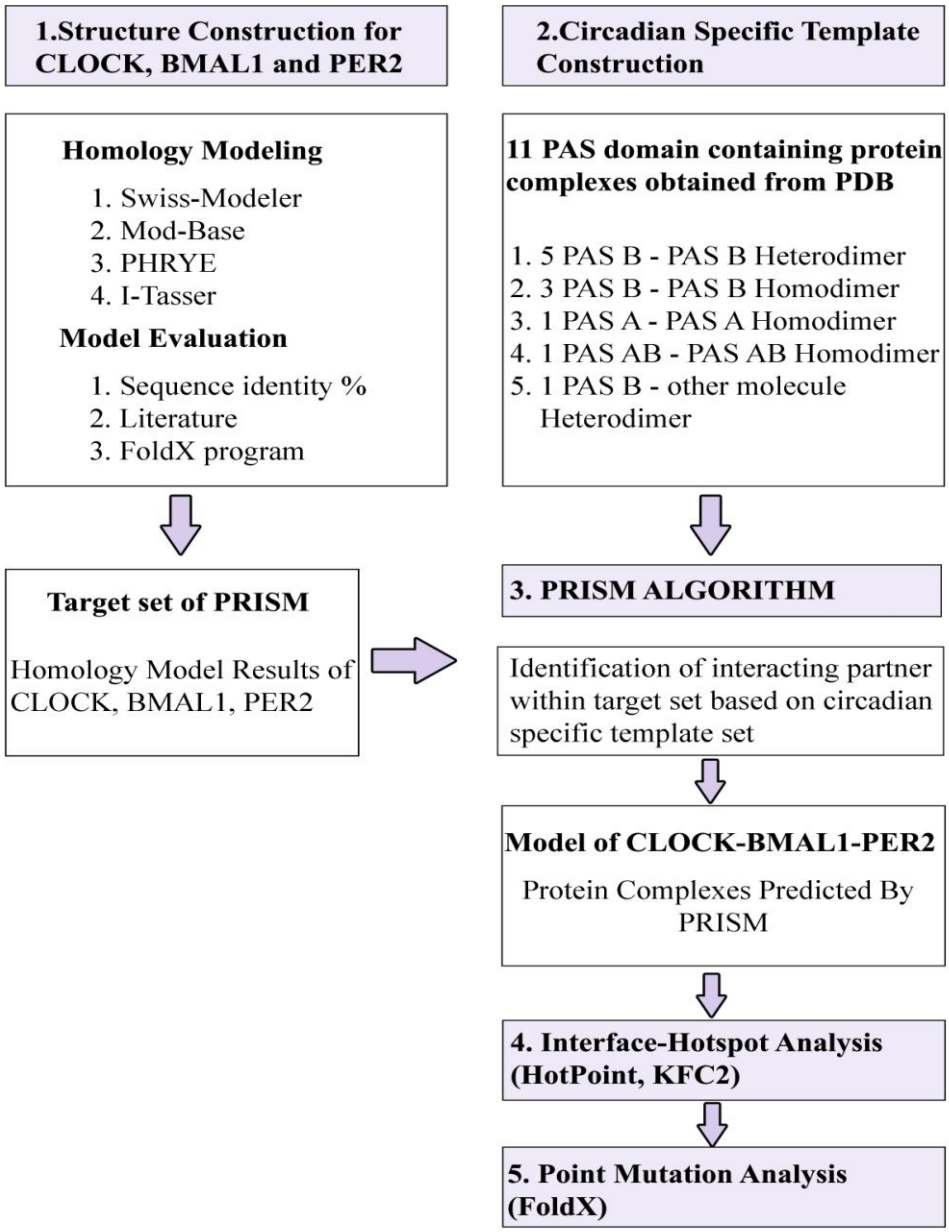


Figure 3-1 Workflow -5 main computational experimental steps performed in this study.(1) Homology modeling, (2) clock pathway specific template construction from PDB (3) high performance prediction algorithm (PRISM), (4) Interface- Hotspot Analysis (HotPoint,KFC2) (5) Mutation Analysis(FoldX)

3.2.1 Homology Modeling

In this work, since there are no available atomic structures for our protein of interests, CLOCK, BMAL1, PER2 and CRY, homology modeling approach is used to obtained 3D structures of proteins. Homology modeling is prediction of the atomic model of proteins on the basis of known experimental structures. Homology modeling is also known as comparative modeling because target protein's model is constructed from its sequence and an experimental three dimensional structure of related homologous protein (template). It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences below a 20% sequence identity can have very different structure. [70].

Different homology modeling programs are used to obtain atomic structures of the BMAL1, PER2, CLOCK and CRY. Those programs are Swiss-model [71]; Modeler-Mode base [72], PHYRE [73], hhpred [74] . The models created by different homology modeling web-servers were evaluated based on 2 restrictions. First one is the sequence identity between target sequence and template structure. The models those have high sequence identity are selected. The quality of the homology model is dependent on the quality of the sequence alignment and template structure. The structures selected according these 2 criteria were selected for further steps for each target in PRISM algorithm.

3.2.2 High Performance Protein-Protein Interaction (PPI) Prediction Algorithm

In this work, we use high performance prediction algorithm PRISM (Protein Interaction by Structural Matching) [64, 75] to analyze multimeric complex formation in clock pathway. PRISM is a web server for the predicting, visualization and analysis

putative protein-protein interactions derived from known protein complex structures in PDB. An interface is the contact region (set of amino acids) between two interacting proteins that links two polypeptide chains by non covalent interaction. This algorithm uses a template interface dataset and a target dataset to predict potential interactions between target dataset proteins based on the template dataset knowledge. The PRISM algorithm searches for whether there are particular surface regions of any two proteins those are spatially similar to the complementary partners of a known interface, in principle these two proteins can interact with each other through these regions.

We used PRISM (Protein Interaction by Structural Matching) [64, 75] to analyze multimeric complex formation in clock pathway. PRISM is a protocol for prediction, visualization and analysis of putative protein-protein interactions. PRISM algorithm uses a template interface dataset and a target dataset to predict potential interactions between target dataset proteins based on the template dataset knowledge An interface is the contact region between two interacting proteins. The template set represents the subset of structurally non-redundant interface architectures. Usage of the known interfaces provides the prediction of new interacting protein pairs. This algorithm searches whether there are particular surface regions of any two proteins those are spatially similar to the complementary partners of a known interface, in principle these two proteins can interact with each other through these regions.

PRISM protocol contains four steps that are processed subsequently. First step is extraction of the surface of the proteins in the target dataset by NACCESS program. Second step is structural alignment; each partner of the template interface is aligned with the target surfaces by Multiprot program. If the matches of the two partner targets contain colliding residues between the putative complexes, they are eliminated. If the putative

complexes passes the residue and hotspot matching thresholds, these targets are transformed on the template interface and form a complex structure. The last step is flexible refinement of the rigid docking solutions of MultiProt to resolve steric clashes in each side chains, and ranking of the complexes by the global energy by usage of FiberDock33, which calculates energies, and ranks the predicted protein complexes.

Template set for PRISM; for this work, we used two different template sets. Default template set (1037-heterodimer set) and case-specific created template set from PDB. Target set contains protein domains of CLOCK, BMAL1 and PER2 separately. To construct this set, all proteins known to be in clock multimeric complex formation and known to have PAS domains are selected which are BMAL1, PER2 and CLOCK (CRY1 and CRY2 is eliminated since they do not have PAS domains). We considered that all these proteins interact via their PAS domains. For that reason, for each protein, only PAS A, only PAS B and PAS AB structures were used as target to analyze interactions on domain level.

3.2.3 Interface Hotspot and Mutation Analysis

Once whole complex of clock proteins was modeled, further analysis were performed on this model to obtain more detailed information about which and how those proteins interact and which residues are critical for those interaction. For this aim three different web servers were used. HotPoint, KFC2 web servers are used for interface and hotspot analysis of protein complexes. Third one is FoldX program, makes energy calculations to obtain quantitative estimation for the importance of residues for the stability of the protein complexes. PositionScan command of FoldX program performs energy calculations to make quantitative estimation for the importance of residues in proteins.

Table 3-1 Clock components specific created template set

PDB id	Interaction Type	Molecule 1	Molecule 2
2A24	pasB-pasB heterodimer	Endothelial PAS domain protein 1	Aryl hydrocarbon receptor nuclear translocator
3A0R	pas-other heterodimer	Sensor protein	Response regulator
3F1N	pasB-pasB heterodimer	Endothelial PAS domain-containing protein 1	Aryl hydrocarbon receptor nuclear translocator
3F1P	pasB-pasB heterodimer	Endothelial PAS domain-containing protein 1	Aryl hydrocarbon receptor nuclear translocator
3H7W	pasB-pasB heterodimer	Endothelial PAS domain-containing protein 1	Aryl hydrocarbon receptor nuclear translocator
3IS2	pas-pas homodimer	Vivid PAS protein VVD	Vivid PAS protein VVD
1WA9	pas AB-pas AB homodimer	Period circadian protein	Period circadian protein
2HV1	pasB-pasB homodimer	Aryl hydrocarbon receptor nuclear translocator	Aryl hydrocarbon receptor nuclear translocator
2P04	pas-pas homodimer	signal transduction histidine kinase	signal transduction histidine kinase
2VLG	pasA-pasA homodimer	Sporulation kinase A	Sporulation kinase A
3F1O	pasB-pasB	Endothelial PAS	Aryl hydrocarbon receptor nuclear

The energy change of protein stability based on mutations that are calculated using FoldX[67] version 3.0 beta 5.1. Each of interface residues in PRISM predictions (the interface between PER2 PAS AB- CLOCK PAS B and between CLOCK PAS B- BMAL1

PAS B) was mutated to other 20 amino acids. The residues those have energy change > 2kcal/mol are considered as critical amino acids for the interactions.

Chapter 4

4 RESULTS

4.1 Structural Models of CLOCK, BMAL1, PER2 and CRY2

The models created by different homology modeling web-servers were evaluated based on two restrictions. The first one is the sequence identity between target sequence and template structure. The second criteria is the regions where models belong on the target sequence- if models do not correspond to PAS domains on the target sequence, they are eliminated, although they have high sequence identities. The evaluated structures of BMAL1, CLOCK and PER2 proteins were selected for further steps.

In CLOCK PAS A domain; Swiss Model provided 26% sequence identity, Modbase provided 13% sequence identity. Although the sequence identities of the models were very different, they resulted in a similar fold, rmsd value is 1.7 Å. Appendix D shows the homology modeling structure comparisons (superimposition and rmsd measurements). Since Swiss Model result in higher sequence identity (26%), it was used as model structure for CLOCK PAS B domain for further analysis. For CLOCK PAS B, two web servers gave high sequence identity. Swiss model's sequence identity is 39% and PHYRE's is 38 %. When their models are superimposed, it is seen that they have similar structure with rmsd value of 2.1 Å. For that reason Swiss model result which is constructed based on template 2h82A is chosen as the model structure.

BMAL1 PAS A domain is modeled by two web servers with close sequence identity. Swiss Model is 26 % and Modbase is 21%. The superimposition of these two models showed that model from Swiss Model and Mode-base have similar structures with rmsd value of 2.4 Å.

Table 4-1 Homology Modeling Results

Protein Name	%Seq id	Template	Residues from-to	Domain	Repository
BMAL1	37.288	1am9C	78-138	Helix-loop-helix motif	SWISSMODEL
	36.036	**1x0oA**	**340-450**	**PAS B**	**SWISSMODEL**
	30	1p97A	347- 447	PAS B	MODBASE
	34	1x0oA	340-447	PAS B	PHYRE
	27.778	**2oolA**	**163-216**	**PAS A**	**SWISSMODEL**
	21	1d06A	152 -215	PAS A	MODBASE
	19.87	**3gdiA**	**147-452**	**PAS AB**	**SWISSMODEL**
CLOCK	**38.532**	**3h82A**	**272-380**	**PAS B**	**SWISSMODEL**
	34	1ll8A	277-378	PAS B	PHYRE
	38.462	1uklD	47-85	HLH motif	SWISSMODEL
	26	**3mjqA**	**114-171**	**PAS A**	**SWISSMODEL**
	13	1d06	108 184	PAS A	MODBASE
	19.636	**1wa9B**	**117-385**	**PAS AB**	**SWISSMODEL**
PER2	**87.829**	**3gdiA**	**170-473**	**PAS AB**	**SWISSMODEL**
	21.667	1wa9B	187-531	PAS AB	SWISSMODEL
CRY2	**51**	**3cvvA**	**4-512**	**-**	**SWISSMODEL**

BMAL1 PAS A domain is modeled by two web servers with close sequence identity. Swiss Model is 26 % and Modbase is 21%. The superimposition of these two models showed that model from Swiss Model and Mode-base has similar structures with rmsd value

of 2.4 Å. So the Swiss Model web server's structure is chosen which is constructed by using 2oolA as template. BMAL1 PAS B domain model is constructed by two web servers that gave high sequence identity.

Swiss model 36% and modbase is 30 %. When their models are superimposed, it is seen that Swiss model and Modbase web servers' models have similar structure with rmsd value of 2.0 Å. PER2 PAS AB domain is modeled by Swiss Model with sequence identity 87% based on template 3gdiA (PER of Drosophila). CRY2 structure is modeled by Swiss Model with sequence identity 51%. Since model quality for PER2 and CRY2 is high, other web servers were not tested.

4.2 Model of the Multimeric Complex

After the modeling of CLOCK, BMAL1, PER2 PAS domains, interaction between these proteins determined using PRISM. PRISM prediction algorithm based on the following concept, if two target proteins contain similar regions to complementary partner of a template interface, it is proposed that these two target proteins interact through these similar complementary regions. The template set is shown in **Table 3-1**. Target set comprises the 7 homology models results which are CLOCK PAS A, CLOCK PAS B, CLOCK PAS AB, BMAL1 PAS A, BMAL1 PAS B, BMAL1 PAS AB and PER2 PAS AB. The PPI predictions of PRISM algorithm is shown in **Table 4-2** .

PRISM predicted interactions between BMAL1- CLOCK and CLCOK -PER2. Totally 34 interactions are found by PRISM algorithm. There is only one interaction (complex) belong to CLOCK-PER2 pair. In **Figure 4-1**, the predicted model for PER2 PAS AB/ CLOCK PAS B complex with its template HIF2-a PAS B/ARNT PAS B complex (3f1nAB) is shown. The other 33 complexes belong to BMAL1-CLOCK pair.

Table 4-2 PRISM results, Totally 34 interactions are found by PRISM algorithm.

Molecule 1	Molecule 2	Template used	No. of interactions
BMAL1 PAS B	CLOCK PAS B	2A24	6
BMAL1PAS B	CLOCK PAS B	3F1P	5
BMAL1 PAS B	CLOCK PAS B	3F1N	1
PER2 PAS AB	CLOCK PAS B	3F1N	1
CLOCK PAS AB	BMAL1 PAS B	3F1N	6
CLOCK PAS AB	BMAL1 PAS B	3F1P	8
CLOCK PAS AB	BMAL1 PAS B	3H7W	7

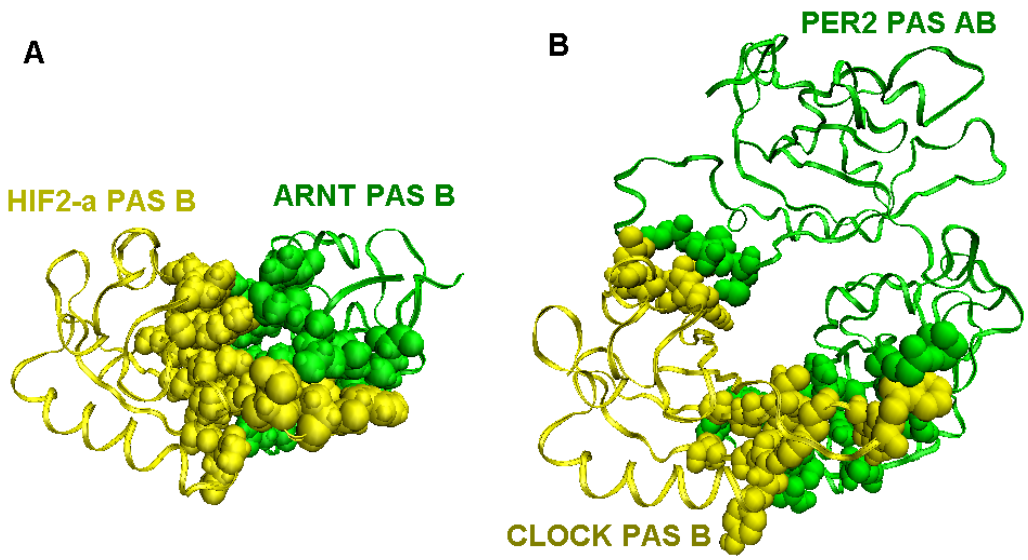


Figure 4-1 The modeled interaction between PER2 PAS AB and CLOCK PAS B domains based on template 3f1n (3f1nA: HIF2-a PAS B, 3f1nB: ARNT PAS B (A) HIF2-a PAS B/ARNT PAS B complex(3f1nAB) (B) Predicted PER2 PAS AB/CLOCK PAS B complex. Interfaces are shown as balls.

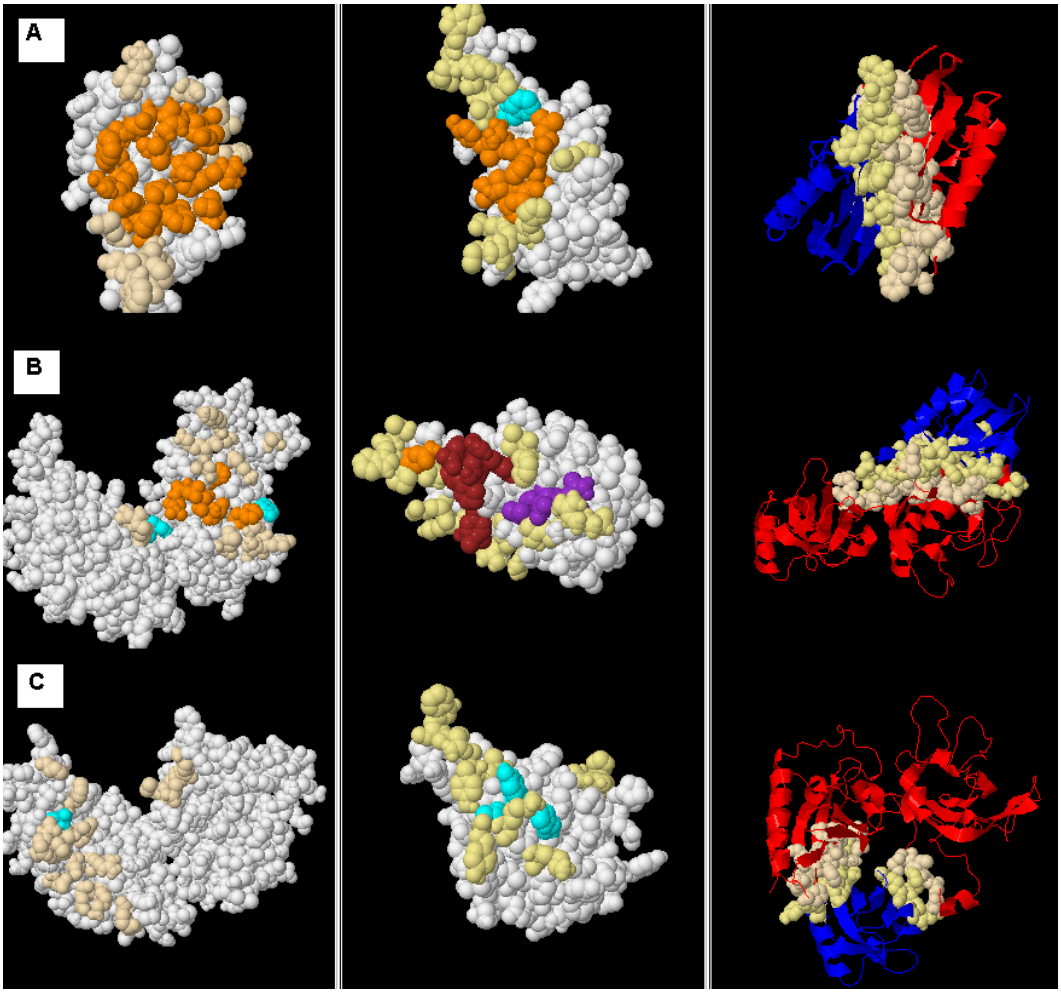


Figure 4-2 Example of Prism results for BMAL1-CLOCK-PER2 protein interactions. Interfaces and hotspots are visualized by using HotPoint web server. The white labeled residues are the interfaces and other colors show the hotspots. Same colored residues imply that those hotspots present in same hot region. (A) CLOCK PAS B- BMAL1 PAS B interaction that is constructed based on the template 2a24 protein complex. Blue chain shows the BMAL1 protein and red chain show the CLOCK protein. (B) CLOCK PAS AB-BMAL1 PAS B interaction which is constructed with the template 3f1n protein complex. Blue chain shows the BMAL1 protein and red chain show the CLOCK protein. (C) PER2 PAS AB – CLOCK PAS B interaction that is predicted by using template 3f1n. Red color represents PER2 and blue color is CLOCK protein.

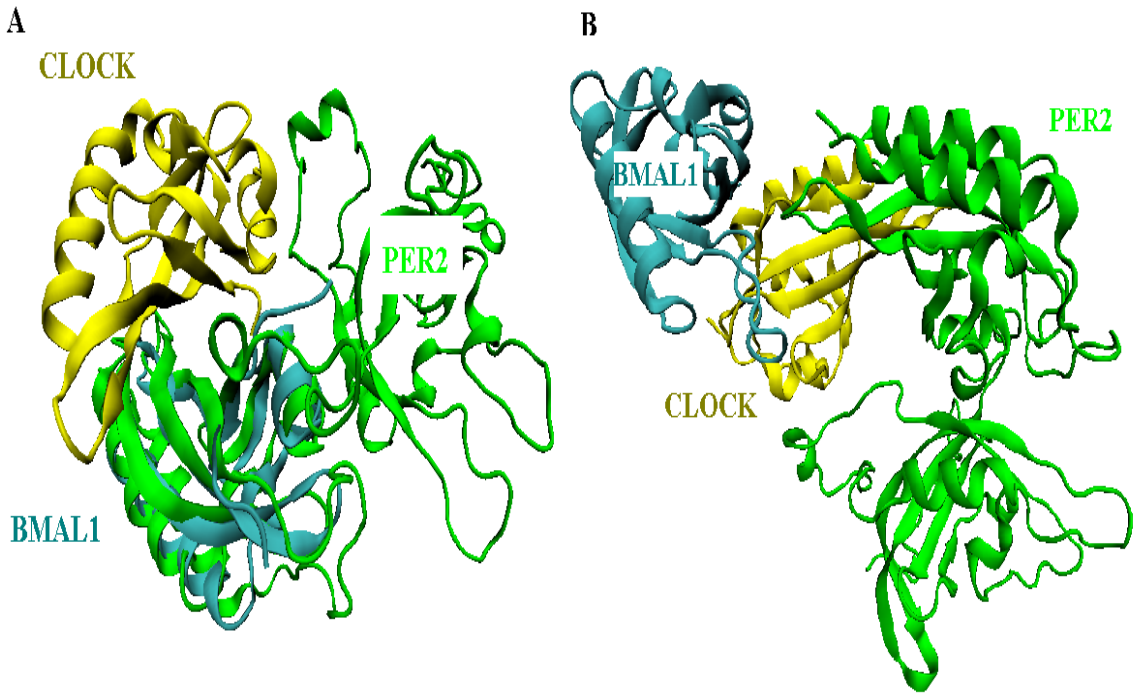


Figure 4-3 Construction of CLOCK-BMAL1-PER2complex model.33 BMAL1 PAS B-CLOCK PAS B complexes were added PER2-CLOCK complex by aligning their CLOCK PAS B regions. (A) One of the eliminated types of CLOCK-BMAL1-PER2 complex is shown. It is seen that BMAL1 PAS B and PER2 PAS B regions are overlapped, totally 9 models obtained like this and they are eliminated. (B) One of the selected the model for CLOCK – BMAL1 –PER2 trimeric complex is shown. It is seen that BMAL1 PAS B and PER2 PAS B uses different interfaces on CLOCK PAS B region. Totally 3 complexes obtained like this and they are selected for further steps.

There are two different types of complexes in 33 predicted interactions. First type of complex is between CLOCK PAS B and BMAL1 PAS B domains, this type of interaction is shown in **Figure 4-2**-B Among the 33 interactions, 12 of the interactions belongs to this type. Second type of interaction is between CLOCK PAS AB and BMAL1 PAS B, the remaining 21 interactions belong to second type, shown in **Figure 4-2**-A. The PPI predictions, obtained from PRISM algorithm were added to each other in a

combinatorial way to obtain CLOCK-BMAL1-PER2 trimeric complex ,using VMD software[76]. Since we have only one complex that contains PER2 protein (interaction predicted between PER2 PAS AB and CLOCK PAS B based on template 3F1N , this PER2-CLOCK complex was used like a skeleton and other complexes were structurally matched on this complex. In structural matching step, we aligned BMAL1 PAS B-CLOCK PAS B complexes (12 complexes) and CLOCK PAS AB and BMAL1 PAS B (21 interactions) on to the PER2 PAS AB- CLOCK PAS B complex by matching their CLOCK PAS B regions. We performed this for all 33 complexes and 30 one them were eliminated since BMAL1 PAS B and PER2 PAS B regions overlapped on CLOCK PAS B interface. Particular interface on a domain can interact with only one partner at the same time, so BMAL1 PAS B and PER2 PAS B have to use different surface patches on CLOCK PAS B domain. The remaining 3 complexes that do not contain overlapping regions were selected as the model for the CLOCK-BMAL1-PER2 trimeric complex.

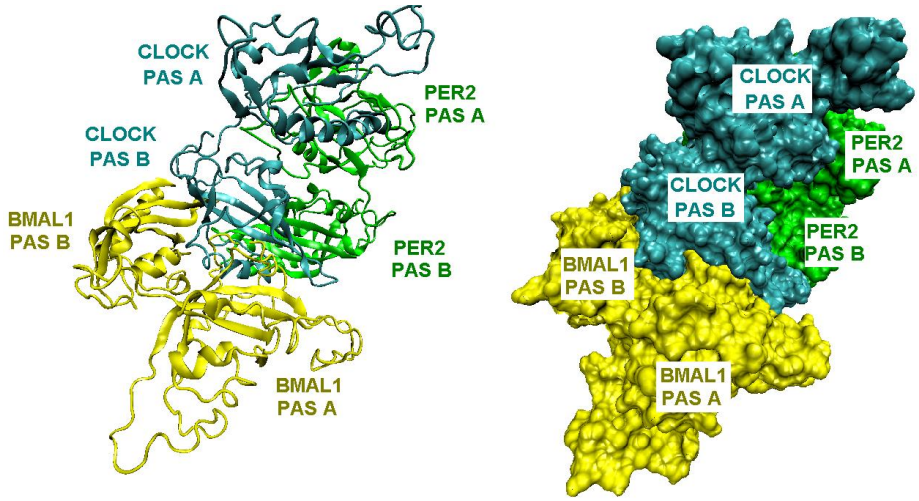


Figure 4-4 Model of CLOCK- BMAL1- PER2 complex. Model was created by combinatorial addition of BMAL1 PAS B-CLOCK PAS B complexes on to the PER2 PAS AB-CLOCK PAS B complex by aligning their CLOCK PASB regions.

4.3 Interface- Hotspot Analysis and FoldX Results

The 3 CLOCK- BMAL1- PER2 complexes that are constructed by putative PPI predictions of PRISM were analyzed with three different web servers, HotPoint, KFC2 and FoldX. We used three different programs to obtain more consistent results by comparing all predictions.

Table 4-3 Interface - hotspot and point mutation analysis results

Interaction Type	Molecule name	Residue No	Residue Type	FoldX Result
PER2- CLOCK	PER2	414	Serine(S)	50 (Histidine)
		429	Arginine (R)	3,6
		431	Lysine (K)	2,0
	CLOCK	332	Glycine(G)	4,5
		354	Histidine(H)	No change
		356	Tyrosine(Y)	1,5
		333	Lysine (K)	2,5
		361	Glutamine(Q)	3,4
CLOCK-BMAL1	CLOCK	317	Aspartic acid (D)	2,5
		338	Tyrosine(Y)	2,7(Glycine)
		350	Tryptophan (W)	3,4
		352	Glutamine(Q)	5,0
		376	Valine(V)	3,5(Proline)
	BMAL1	347	Valine(V)	No change
		349	Arginine (R)	2,7
		362	Arginine(R)	3,5(Tyrosine)
		405	Isoleucine(I)	2,0
		427	Phenylalanine(F)	5,0(Proline)
		429	Phenylalanine(F)	3,5
		441	Valine(V)	4,5(Histidine)

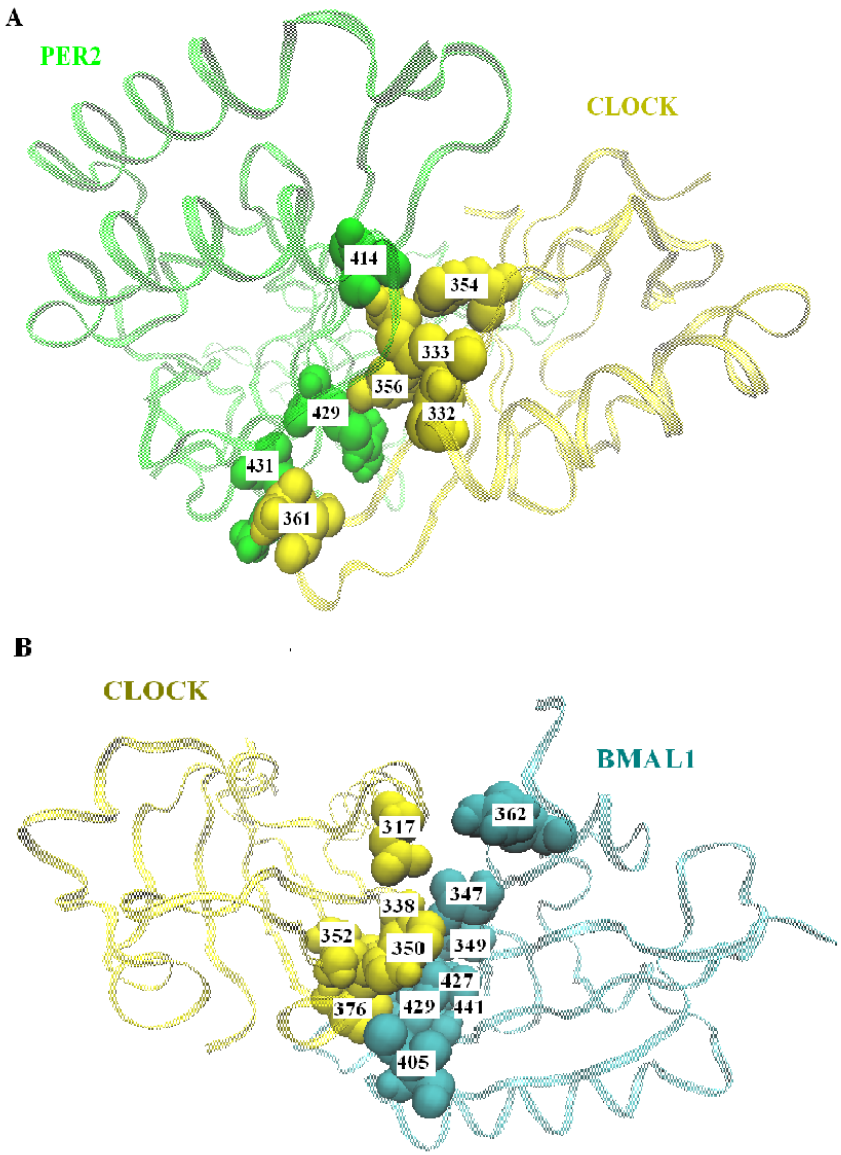


Figure 4-5 Hotspots predicted by HotPoint and KFC2 web servers.The molecules PER2, CLOCK, BMAL1 are colored green, yellow and cyan. There two interfaces present in this complex, first one is between PER2 and CLOCK and second one is between CLOCK and BMAL1. (A) The hotspots on the interface of PER2-CLOCK are shown. (B) The hotspots on the interface of CLOCK-BMAL1 are shown.

In CLOCK-BMAL1-PER2 complex, there are 2 interactions. First one is between PER2 and CLOCK proteins and second one is between BMAL1 and CLOCK proteins. Each interaction was analyzed separately using HotPoint and KFC2 web servers. Each of the protein stability was analyzed using FoldX program to validate HotPoint predictions. Results are shown in **Table** 4-3 Interface - hotspot **and point mutation** analysis results. Hot spot results were validated by FoldX program.e.g. 332. residue, glycine on CLOCK protein is identified as hotspot by HotPoint server. In FoldX results, it is also seen that when it is mutated to valine, energy of the protein increases 4.5 kcal/mol, that confirms the result of the HotPoint server. These analysis provide us information regarding to residues those are critical for the interactions between clock proteins. In **Figure 4-5**, the hotspots, labeled on complex those predicted by at least two of three programs are shown. In CLOCK-PER2 complex, it is seen that anti- parallel β-sheet region of PAS B domains of CLOCK and PER2 are used as interface. Besides, some residues on alpha helix residues of PER2 PAS A are found as hotspots by HotPoint. In BMAL1- CLOCK complex, it is seen that the beta-sheet regions on PAS B domains are used.

Chapter 5

5 DISCUSSION

In this work, we first sought to identify interactions within each of the clock proteins that have PAS domains, using high performance structure based prediction algorithm, PRISM. PAS-PAS domain containing protein complexes, available in PDB were used as template for PPI prediction. Homology models of CLOCK, BMAL1 and PER2 were used as target for PPI prediction. The prediction algorithm based on the following concept: if two target proteins contain similar regions to complementary partner of a template interface, it is assumed that these two target proteins interact through these similar complementary regions. PAS domain containing clock proteins' trimeric complex, CLOCK-BMAL1-PER2 was constructed using PRISM algorithm. Each of the interactions between two individual proteins that we found and previous findings (**Appendix C)** about them will be discussed below.

We found that CLOCK and BMAL1 form a heterodimer via their PAS B domains. The CLOCK-BMAL1 interaction was predicted with PRISM by using 4 different templates, HIF2-a PAS B/ARNT PAS B complex (3f1nAB) , C-terminal PAS domain (PAS-B)/ Aryl hydrocarbon receptor nuclear translocator (2a24AB), HIF2 alpha C-terminal PAS domain/Aryl hydrocarbon receptor nuclear translocator (3f1pAB), HIF2alpha C-terminal PAS domain/ Aryl hydrocarbon receptor nuclear translocator (3h7wAB). Then, PAS A domains of CLOCK and BMAL1 were added to CLOCK PAS B- BMAL1 PAS B protein complex by using VMD. When we align PAS A domain on this complex, we observed that PAS A domains do not oriented in parallel to each other, they placed on different axis.(see **Figure 5-1**). According to interface and hotspot analysis, it is seen that the beta-sheet regions on PAS B domain are used as interface.

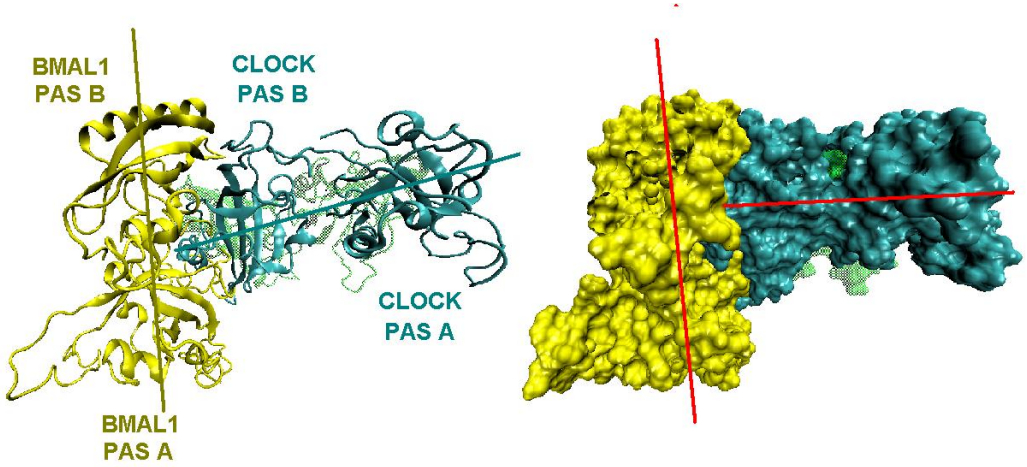


Figure 5-1 The view of CLOCK PAS AB and BMAL1 PAS AB

In our results, we obtained two different types of CLOCK- BMAL1 heterodimer, CLOCK PAS B- BMAL1 PAS B and CLOCK PAS AB- BMAL1 PAS B. According to the complex energy calculation of PRISM, first type of interactions have energies approximately -20 kcal/mol and second type interactions CLOCK PAS AB-BMAL1 PAS B interactions have energies approximately -40kcal/mol .(Appendix B shows the energy of complexes those predicted by PRISM). These results imply that, PAS A domain of CLOCK has additive effect on CLOCK-BMAL1 PAS B -PAS B heterodimerization. In a previous study, the importance of PAS domains for BMAL1-CLOCK heterodimerization was tested and it was seen that single PAS domain cannot sustain heterodimerization alone (Appendix C, [60, 77]). Both PAS AB domains from BMAL1 and CLOCK are required for heterodimerization. In addition to this, it is seen that PAS B domain of BMAL1 can contribute interaction with CLOCK PAS AB [9]. Second type of PRISM interaction prediction; support the findings of this study since CLOCK PAS AB-BMAL1 PAS B complexes have lower energies than CLOCK PAS B- BMAL1 PAS B complexes. As a result, our model propose that PAS A domains of CLOCK and BMAL1 are not used as

interface for heterodimerization but PAS A domains have additive influence for PAS B-PAS B heterodimer formation.

Until now, the interaction between CLOCK and PER2 is shown in many studies [21, 60, 62, 63]. However, the interaction regions within those proteins those are required for binding are not studied neither at domain level nor residual level. According to our prediction algorithm, CLOCK interacts with PER2, which holds with the previous findings. . In our methodology, one CLOCK-PER2 protein complex was constructed based on template HIF2-a PAS B/ARNT PAS B complex (3f1nAB). In this complex, PAS B domain of CLOCK and PAS B domain of PER2 are used as interface. Besides, some residues on alpha helix region of PER2 PAS A are found as hotspots by HotPoint.

The interaction between BMAL1 and PER2 was observed by using mammalian two hybrid system [21, 62]. Additionally, it is found that PER2 binds to PAS A domain of BMAL1 [60]. In our system, we did not find any interaction between BMAL1 and PER2 in opposite to previous findings. However another study observed that, when CRY proteins are added this complex, CRY1 and CRY2 influence the interaction between PER2 and BMAL1. The CRY proteins moreover seem to have a higher affinity to BMAL1 than PER2. When they added PER2 to BMAL-CRY complex, it is seen that, PER2 does not significantly affect the interaction between CRY and BMAL1 [9]. All these findings propose that, PER2 do not interact with BMAL1 when they are found as the multimeric complex of clock proteins, CLOCK-BMAL1-PER2- CRY2. Although binary interactions between PER2 and BMAL1 were experimentally shown, their interaction is needed to be validated when PER2 and BMAL1 are present in the multimeric complex. The cartoon representation of whole complex is presented in **Figure 5-2**.

In conclusion, to complete our aim and characterize the interactions of BMAL1, CLOCK and PER2, we use PRISM server. Via PRISM protocol, we are able to predict which interactions can and cannot co-exist. Here we apply this concept to clock proteins, BMAL1, CLOCK and PER2. We predict, BMAL1 and CLOCK interacts with each other through their PAS B domains and the PER2 interacts with this dimer through the PAS B domain of the CLOCK. In BMAL1/CLOCK interface, we found 12 hotspots, 7 residues on BMAL1 (347,349,362,405,427,429,441), 5 residues on CLOCK (317, 338, 350, 352,376). In PER2/CLOCK interface, 8 hotspots are found, 3 residues on PER2 (414, 429,431) and 5 residues on CLOCK (332,354,356,333,361). We believe that such a strategy should be very useful in the actual comprehension of PPI. For experimental part, 20 hotspots were found, shown in **Table 4-3**. These hotspots are also needed to be validated by mutagenesis analysis

As a future work, for the modeling part, generation of structures which leads the improvements in the template set is crucial for computational analysis part of this study. In our results, we show that PAS B domains of CLOCK and BMAL1 are used as interface for heterodimerization. However, the role of PAS A domains for heterodimerization are still needed to be searched.

A better understanding of the interactions between clock proteins will clearly have an additive effect on our understanding of other diseases and human health. There is a direct link between circadian rhythms and metabolic gene regulation, as well as nutrient uptake. So the mechanisms by which circadian rhythms are maintained are very essential to elucidate the cross-talk with metabolic signaling. These studies- analysis of clock regulation is not only important to understand clock mechanism at molecular level but also

will lead to improve novel therapeutic approaches in the future against clock-regulated diseases, like Jet-Lag.

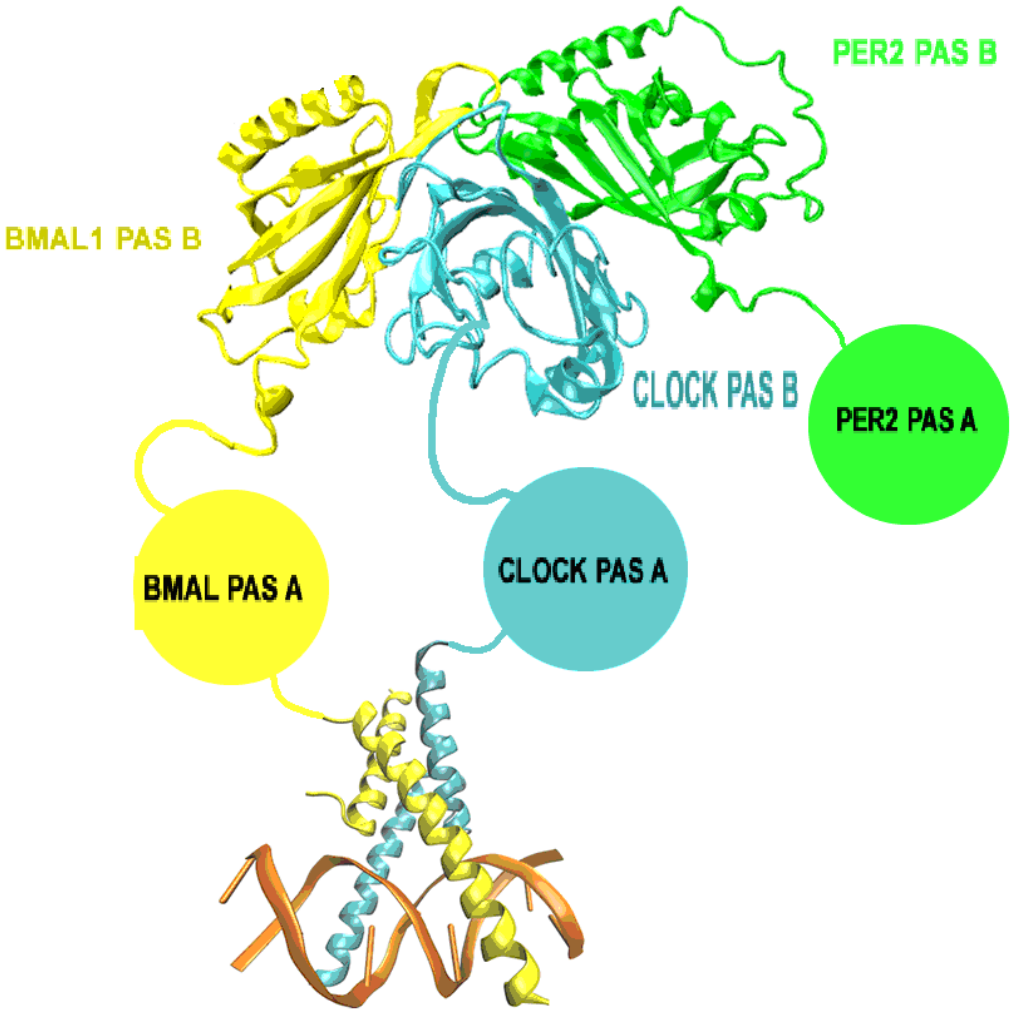
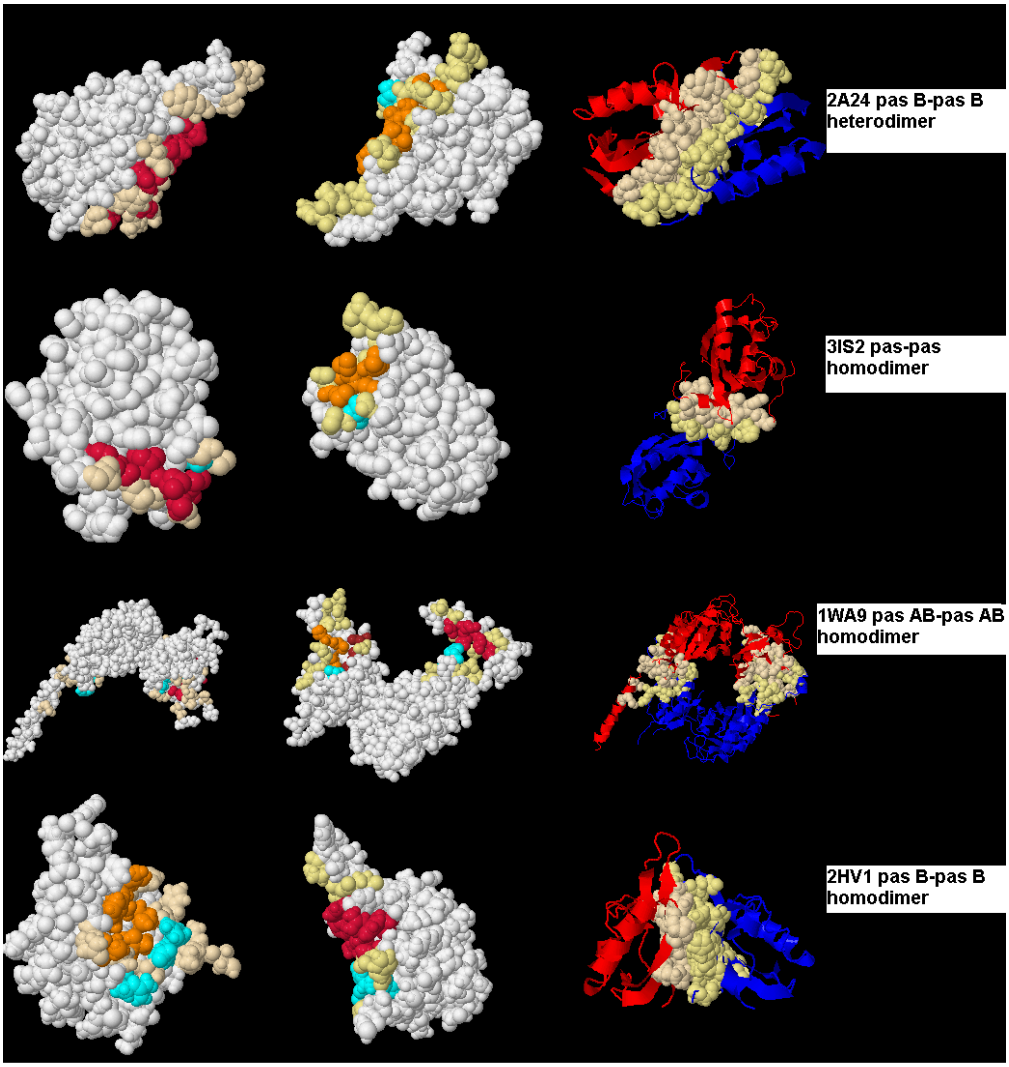


Figure 5-2 The cartoon representation of BMAL1-CLOCK-PER2 complex model. Proteins are labeled on model; the interactions predicted by PRISM are illustrated here by using new cartoon representation. Some parts of molecules are represented as balls, because the interactions in these parts are not clarified.

6 Appendix

6.1 Appendix A: 4 different type of interactions within template set



6.2 Appendix B: PRISM Results

Template	Mol 1	Mol2	Energy (kcal/mol)
2a24AB	BMAL1 PAS B	CLOCK PAS AB	-10.15
2a24AB	BMAL1 PAS B	CLOCK PAS AB	-15.08
2a24AB	CLOCK PAS B	BMAL1 PAS B	-14.33
2a24AB	CLOCK PAS B	BMAL1 PAS B	-14.33
2a24AB	CLOCK PAS B	BMAL1 PAS B	-21.32
3f1pAB	CLOCK PAS B	BMAL1 PAS B	-19.95
2a24AB	CLOCK PAS B	BMAL1 PAS B	-11.32
3f1pAB	CLOCK PAS B	BMAL1 PAS B	-9.95
3f1pAB	CLOCK PAS B	BMAL1 PAS B	-9.61
3f1pAB	CLOCK PAS B	BMAL1 PAS B	-19.63
3f1pAB	CLOCK PAS B	BMAL1 PAS B	-21.69
3f1pAB	BMAL1 PAS B	CLOCK PAS AB	-13.04
3f1pAB	BMAL1 PAS B	CLOCK PAS B	-12.2
3f1pAB	BMAL1 PAS B	CLOCK PAS B	-11.52
3f1pAB	CLOCK PAS AB	BMAL1 PAS B	-34.05
3f1pAB	CLOCK PAS AB	BMAL1 PAS B	-42.7

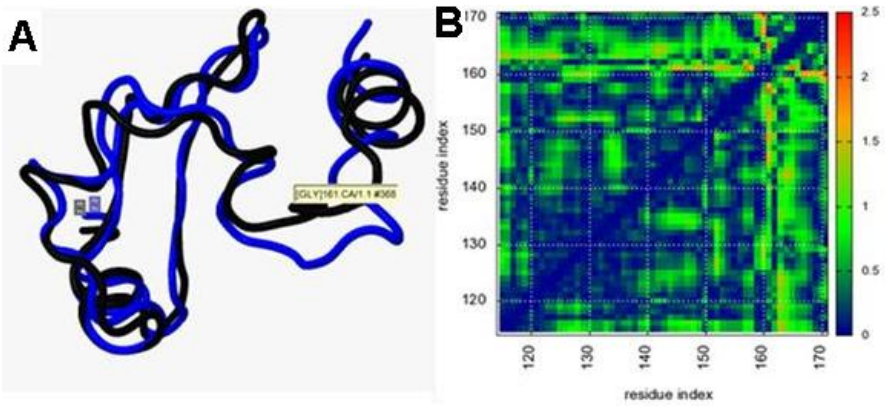
3f1pAB	CLOCK PAS AB	BMAL1 PAS B	-24.72
3f1pAB	CLOCK PAS AB	BMAL1 PAS B	-22.41
3f1pAB	CLOCK PAS AB	BMAL1 PAS B	-21.52
3f1pAB	CLOCK PAS AB	BMAL1 PAS B	-42.41
3f1pAB	CLOCK PAS AB	BMAL1 PAS B	-34.95
3f1pAB	CLOCK PAS AB	BMAL1 PAS B	-40.7
3f1pAB	CLOCK PAS AB	BMAL1 PAS B	-31.5
3f1pAB	CLOCK PAS AB	BMAL1 PAS B	-36.53
3f1pAB	CLOCK PAS AB	BMAL1 PAS B	-39.53
3h7wAB	CLOCK PAS B	BMAL1 PAS B	-9.6
3h7wAB	CLOCK PAS AB	BMAL1 PAS B	-42.22
3h7wAB	CLOCK PAS AB	BMAL1 PAS B	-43.2
3h7wAB	CLOCK PAS AB	BMAL1 PAS B	-35.93
3h7wAB	CLOCK PAS AB	BMAL1 PAS B	-58.91
3h7wAB	CLOCK PAS AB	BMAL1 PAS B	-39.7

6.3 **Appendix C: Experimentally Validated Protein Interactions within clock proteins** (The IntAct molecular interaction database is used)

Protein Name	**Protein Name**	**Interaction Detection Method**	**Species**	**Interaction Domain**	**Complex Type**	**Ref.**
BMAL1	CLOCK	Two-hybrid	Homo Sapiens	Both PAS A and PAS B from BMAL1 and CLOCK is required.	DNA-off Binary interaction	[60, 77]
BMAL1	CRY1	gal4 vp16 complementation	Homo Sapiens	C-terminal of BMAL1 is required.	DNA-off Binary interaction	[60]
		anti bait coimmunoprecipitation	Mus musculus		DNA-off Binary interaction	[78]
BMAL1	PER2	gal4 vp16 complementation	Homo Sapiens	PAS B of BMAL 1 is essential.	DNA-off Binary interaction	[60]
CLOCK	PER2	anti bait coimmunoprecipitation	Mus musculus	Not clarified.	DNA-off Binary interaction	[63]
		gal4 vp16 complementation	Homo Sapiens		DNA-off Binary interaction	[60]
CLOCK	CRY2	Two-hybrid	Mus musculus	Not clarified.	DNA-off Binary interaction	[22]
CRY2	BMAL1+CLOCK+ E -BOX COMPLEX	ChIP experiments	Mus musculus	Not clarified.	DNA-on interaction	[56]

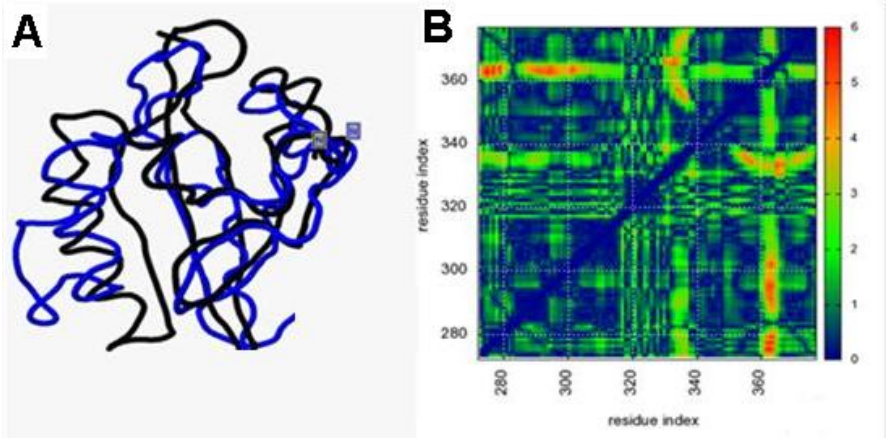
6.4 Appendix D: Homology Modeling Structure Comparison Results

CLOCK PAS A



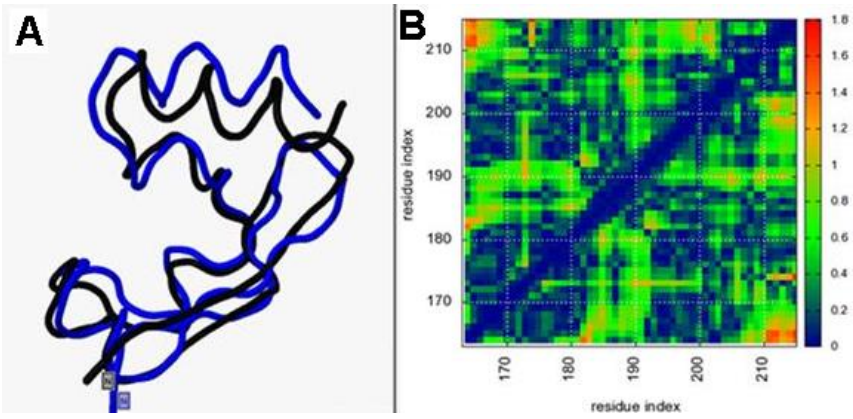
(A)Superimposition of the models, Blue-swiss model, Black-modbase model, RMSD is 1.7 A. (B) Degree of variability within residues of models

CLOCK PAS B



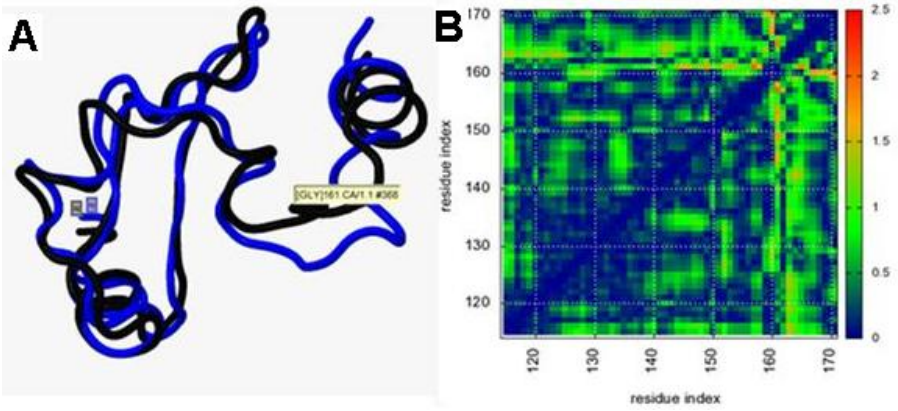
(A)Superimposition of the models, Blue-swiss model, Black-PHYRE model, RMSD is 2.1 A. (B) Degree of variability within residues of models

BMAL1 PAS A



(A)Superimposition of the models, Blue-swiss model, Black-modbase model, RMSD is 2.4 A. (B) Degree of variability within residues of models

BMAL1 PAS B



(A)Superimposition of the models, Blue-swiss model, Black-modbase model, RMSD is 2.0 A. (B) Degree of variability within residues of models

7 Bibliography

1. Ouyang Y, Andersson CR, Kondo T, Golden SS, Johnson CH: **Resonating circadian clocks enhance fitness in cyanobacteria**. *Proc Natl Acad Sci U S A* 1998, **95**(15):8660-8664.
2. Dodd AN, Salathia N, Hall A, Kevei E, Toth R, Nagy F, Hibberd JM, Millar AJ, Webb AA: **Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage**. *Science* 2005, **309**(5734):630-633.
3. Gekakis N, Staknis D, Nguyen HB, Davis FC, Wilsbacher LD, King DP, Takahashi JS, Weitz CJ: **Role of the CLOCK protein in the mammalian circadian mechanism**. *Science* 1998, **280**(5369):1564-1569.
4. Travnickova-Bendova Z, Cermakian N, Reppert SM, Sassone-Corsi P: **Bimodal regulation of mPeriod promoters by CREB-dependent signaling and CLOCK/BMAL1 activity**. *Proc Natl Acad Sci U S A* 2002, **99**(11):7728-7733.
5. Yoo SH, Ko CH, Lowrey PL, Buhr ED, Song EJ, Chang S, Yoo OJ, Yamazaki S, Lee C, Takahashi JS: **A noncanonical E-box enhancer drives mouse Period2 circadian oscillations in vivo**. *Proc Natl Acad Sci U S A* 2005, **102**(7):2608-2613.
6. Griffin EA, Jr., Staknis D, Weitz CJ: **Light-independent role of CRY1 and CRY2 in the mammalian circadian clock**. *Science* 1999, **286**(5440):768-771.
7. Kume K, Zylka MJ, Sriram S, Shearman LP, Weaver DR, Jin X, Maywood ES, Hastings MH, Reppert SM: **mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop**. *Cell* 1999, **98**(2):193-205.
8. Chaves I, Yagita K, Barnhoorn S, Okamura H, van der Horst GT, Tamanini F: **Functional evolution of the photolyase/cryptochrome protein family: importance of the C terminus of mammalian CRY1 for circadian core oscillator performance**. *Mol Cell Biol* 2006, **26**(5):1743-1753.
9. Dardente H, Fortier EE, Martineau V, Cermakian N: **Cryptochromes impair phosphorylation of transcriptional activators in the clock: a general mechanism for circadian repression**. *Biochem J* 2007, **402**(3):525-536.
10. Kiyohara YB, Tagao S, Tamanini F, Morita A, Sugisawa Y, Yasuda M, Yamanaka I, Ueda HR, van der Horst GT, Kondo T *et al*: **The BMAL1 C terminus regulates the circadian transcription feedback loop**. *Proc Natl Acad Sci U S A* 2006, **103**(26):10074-10079.
11. Kondratov RV, Kondratova AA, Lee C, Gorbacheva VY, Chernov MV, Antoch MP: **Post-translational regulation of circadian transcriptional CLOCK(NPAS2)/BMAL1 complex by CRYPTOCHROMES**. *Cell Cycle* 2006, **5**(8):890-895.

12. Giebultowicz J: **Chronobiology: biological timekeeping**. *Integr Comp Biol* 2004, **44**(3):266.
13. Reppert SM, Weaver DR: **Molecular analysis of mammalian circadian rhythms**. *Annu Rev Physiol* 2001, **63**:647-676.
14. Jomary C, Jones SE, Lotery AJ: **Generation of light-sensitive photoreceptor phenotypes by genetic modification of human adult ocular stem cells with Crx**. *Invest Ophthalmol Vis Sci* 2010, **51**(2):1181-1189.
15. King DP, Takahashi JS: **Molecular genetics of circadian rhythms in mammals**. *Annu Rev Neurosci* 2000, **23**:713-742.
16. Harmer SL, Panda S, Kay SA: **Molecular bases of circadian rhythms**. *Annu Rev Cell Dev Biol* 2001, **17**:215-253.
17. Nyce J, Binkley S: **Extraretinal photoreception in chickens: entrainment of the circadian locomotor activity rhythm**. *Photochem Photobiol* 1977, **25**(6):529-531.
18. Berson DM, Dunn FA, Takao M: **Phototransduction by retinal ganglion cells that set the circadian clock**. *Science* 2002, **295**(5557):1070-1073.
19. Lehman MN, Silver R, Gladstone WR, Kahn RM, Gibson M, Bittman EL: **Circadian rhythmicity restored by neural transplant. Immunocytochemical characterization of the graft and its integration with the host brain**. *J Neurosci* 1987, **7**(6):1626-1638.
20. Lee C, Etchegaray JP, Cagampang FR, Loudon AS, Reppert SM: **Posttranslational mechanisms regulate the mammalian circadian clock**. *Cell* 2001, **107**(7):855-867.
21. Lee C, Weaver DR, Reppert SM: **Direct association between mouse PERIOD and CKIepsilon is critical for a functioning circadian clock**. *Mol Cell Biol* 2004, **24**(2):584-594.
22. Shearman LP, Sriram S, Weaver DR, Maywood ES, Chaves I, Zheng B, Kume K, Lee CC, van der Horst GT, Hastings MH *et al*: **Interacting molecular loops in the mammalian circadian clock**. *Science* 2000, **288**(5468):1013-1019.
23. Reddy P, Zehring WA, Wheeler DA, Pirrotta V, Hadfield C, Hall JC, Rosbash M: **Molecular analysis of the period locus in Drosophila melanogaster and identification of a transcript involved in biological rhythms**. *Cell* 1984, **38**(3):701-710.
24. Hardin PE, Hall JC, Rosbash M: **Feedback of the Drosophila period gene product on circadian cycling of its messenger RNA levels**. *Nature* 1990, **343**(6258):536-540.
25. Vanselow K, Vanselow JT, Westermark PO, Reischl S, Maier B, Korte T, Herrmann A, Herzel H, Schlosser A, Kramer A: **Differential effects of PER2**

phosphorylation: molecular basis for the human familial advanced sleep phase syndrome (FASPS). *Genes Dev* 2006, **20**(19):2660-2672.

26. Vitaterna MH, King DP, Chang AM, Kornhauser JM, Lowrey PL, McDonald JD, Dove WF, Pinto LH, Turek FW, Takahashi JS: **Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior**. *Science* 1994, **264**(5159):719-725.
27. Antoch MP, Song EJ, Chang AM, Vitaterna MH, Zhao Y, Wilsbacher LD, Sangoram AM, King DP, Pinto LH, Takahashi JS: **Functional identification of the mouse circadian Clock gene by transgenic BAC rescue**. *Cell* 1997, **89**(4):655-667.
28. King DP, Zhao Y, Sangoram AM, Wilsbacher LD, Tanaka M, Antoch MP, Steeves TD, Vitaterna MH, Kornhauser JM, Lowrey PL *et al*: **Positional cloning of the mouse circadian clock gene**. *Cell* 1997, **89**(4):641-653.
29. Debruyne JP, Noton E, Lambert CM, Maywood ES, Weaver DR, Reppert SM: **A clock shock: mouse CLOCK is not required for circadian oscillator function**. *Neuron* 2006, **50**(3):465-477.
30. DeBruyne JP, Weaver DR, Reppert SM: **CLOCK and NPAS2 have overlapping roles in the suprachiasmatic circadian clock**. *Nat Neurosci* 2007, **10**(5):543-545.
31. Doi M, Hirayama J, Sassone-Corsi P: **Circadian regulator CLOCK is a histone acetyltransferase**. *Cell* 2006, **125**(3):497-508.
32. Hogenesch JB, Gu YZ, Jain S, Bradfield CA: **The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors**. *Proc Natl Acad Sci U S A* 1998, **95**(10):5474-5479.
33. Tsutsui K, Maeda Y, Tsutsui K, Seki S, Tokunaga A: **cDNA cloning of a novel amphiphysin isoform and tissue-specific expression of its multiple splice variants**. *Biochem Biophys Res Commun* 1997, **236**(1):178-183.
34. Bunger MK, Wilsbacher LD, Moran SM, Clendenin C, Radcliffe LA, Hogenesch JB, Simon MC, Takahashi JS, Bradfield CA: **Mop3 is an essential component of the master circadian pacemaker in mammals**. *Cell* 2000, **103**(7):1009-1017.
35. McDearmon EL, Patel KN, Ko CH, Walisser JA, Schook AC, Chong JL, Wilsbacher LD, Song EJ, Hong HK, Bradfield CA *et al*: **Dissecting the functions of the mammalian clock protein BMAL1 by tissue-specific rescue in mice**. *Science* 2006, **314**(5803):1304-1308.
36. Preitner N, Damiola F, Lopez-Molina L, Zakany J, Duboule D, Albrecht U, Schibler U: **The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator**. *Cell* 2002, **110**(2):251-260.

37. Guillaumond F, Dardente H, Giguere V, Cermakian N: **Differential control of Bmal1 circadian transcription by REV-ERB and ROR nuclear receptors**. *J Biol Rhythms* 2005, **20**(5):391-403.
38. Sato TK, Panda S, Miraglia LJ, Reyes TM, Rudic RD, McNamara P, Naik KA, FitzGerald GA, Kay SA, Hogenesch JB: **A functional genomics strategy reveals Rora as a component of the mammalian circadian clock**. *Neuron* 2004, **43**(4):527-537.
39. Liu AC, Tran HG, Zhang EE, Priest AA, Welsh DK, Kay SA: **Redundant function of REV-ERBalpha and beta and non-essential role for Bmal1 cycling in transcriptional regulation of intracellular circadian rhythms**. *PLoS Genet* 2008, **4**(2):e1000023.
40. Sun ZS, Albrecht U, Zhuchenko O, Bailey J, Eichele G, Lee CC: **RIGUI, a putative mammalian ortholog of the Drosophila period gene**. *Cell* 1997, **90**(6):1003-1011.
41. Zheng B, Larkin DW, Albrecht U, Sun ZS, Sage M, Eichele G, Lee CC, Bradley A: **The mPer2 gene encodes a functional component of the mammalian circadian clock**. *Nature* 1999, **400**(6740):169-173.
42. Miyazaki K, Mesaki M, Ishida N: **Nuclear entry mechanism of rat PER2 (rPER2): role of rPER2 in nuclear localization of CRY protein**. *Mol Cell Biol* 2001, **21**(19):6651-6659.
43. Zheng B, Albrecht U, Kaasik K, Sage M, Lu W, Vaishnav S, Li Q, Sun ZS, Eichele G, Bradley A *et al*: **Nonredundant roles of the mPer1 and mPer2 genes in the mammalian circadian clock**. *Cell* 2001, **105**(5):683-694.
44. Reppert SM, Weaver DR: **Coordination of circadian timing in mammals**. *Nature* 2002, **418**(6901):935-941.
45. Taylor BL, Zhulin IB: **PAS domains: internal sensors of oxygen, redox potential, and light**. *Microbiol Mol Biol Rev* 1999, **63**(2):479-506.
46. Hefti MH, Francoijs KJ, de Vries SC, Dixon R, Vervoort J: **The PAS fold. A redefinition of the PAS domain based upon structural prediction**. *Eur J Biochem* 2004, **271**(6):1198-1208.
47. McIntosh BE, Hogenesch JB, Bradfield CA: **Mammalian Per-Arnt-Sim proteins in environmental adaptation**. *Annu Rev Physiol, 2010*, **72**:625-645.
48. Moglich A, Ayers RA, Moffat K: **Structure and signaling mechanism of Per-ARNT-Sim domains**. *Structure* 2009, **17**(10):1282-1294.
49. Hogenesch JB, Chan WK, Jackiw VH, Brown RC, Gu YZ, Pray-Grant M, Perdew GH, Bradfield CA: **Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway**. *J Biol Chem* 1997, **272**(13):8581-8593.

50. Zhou YD, Barnard M, Tian H, Li X, Ring HZ, Francke U, Shelton J, Richardson J, Russell DW, McKnight SL: **Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system**. *Proc Natl Acad Sci U S A* 1997, **94**(2):713-718.
51. Yang J, Zhang L, Erbel PJ, Gardner KH, Ding K, Garcia JA, Bruick RK: **Functions of the Per/ARNT/Sim domains of the hypoxia-inducible factor**. *J Biol Chem* 2005, **280**(43):36047-36054.
52. Yildiz O, Doi M, Yujnovsky I, Cardone L, Berndt A, Hennig S, Schulze S, Urbanke C, Sassone-Corsi P, Wolf E: **Crystal structure and interactions of the PAS repeat region of the Drosophila clock protein PERIOD**. *Mol Cell* 2005, **17**(1):69-82.
53. Miyatake H, Mukai M, Park SY, Adachi S, Tamura K, Nakamura H, Nakamura K, Tsuchiya T, Iizuka T, Shiro Y: **Sensory mechanism of oxygen sensor FixL from Rhizobium meliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies**. *J Mol Biol* 2000, **301**(2):415-431.
54. Kurokawa H, Lee DS, Watanabe M, Sagami I, Mikami B, Raman CS, Shimizu T: **A redox-controlled molecular switch revealed by the crystal structure of a bacterial heme PAS sensor**. *J Biol Chem* 2004, **279**(19):20186-20193.
55. Card PB, Erbel PJ, Gardner KH: **Structural basis of ARNT PAS-B dimerization: use of a common beta-sheet interface for hetero- and homodimerization**. *J Mol Biol* 2005, **353**(3):664-677.
56. Ye R, Selby CP, Ozturk N, Annayev Y, Sancar A: **Biochemical analysis of the canonical model for the Mammalian circadian clock**. *J Biol Chem 2011*, **286**(29):25891-25902.
57. Darlington TK, Wager-Smith K, Ceriani MF, Staknis D, Gekakis N, Steeves TD, Weitz CJ, Takahashi JS, Kay SA: **Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim**. *Science* 1998, **280**(5369):1599-1603.
58. Hao H, Allen DL, Hardin PE: **A circadian enhancer mediates PER-dependent mRNA cycling in Drosophila melanogaster**. *Mol Cell Biol* 1997, **17**(7):3687-3693.
59. Rutila JE, Suri V, Le M, So WV, Rosbash M, Hall JC: **CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of Drosophila period and timeless**. *Cell* 1998, **93**(5):805-814.
60. Langmesser S, Tallone T, Bordon A, Rusconi S, Albrecht U: **Interaction of circadian clock proteins PER2 and CRY with BMAL1 and CLOCK**. *BMC Mol Biol* 2008, **9**:41.

61. Ozber N, Baris I, Tatlici G, Gur I, Kilinc S, Unal EB, Kavakli IH: **Identification of two amino acids in the C-terminal domain of mouse CRY2 essential for PER2 interaction**. *BMC Mol Biol* 2010, **11**:69.
62. Etchegaray JP, Yang X, DeBruyne JP, Peters AH, Weaver DR, Jenuwein T, Reppert SM: **The polycomb group protein EZH2 is required for mammalian circadian clock function**. *J Biol Chem* 2006, **281**(30):21209-21215.
63. Asher G, Gatfield D, Stratmann M, Reinke H, Dibner C, Kreppel F, Mostoslavsky R, Alt FW, Schibler U: **SIRT1 regulates circadian clock gene expression through PER2 deacetylation**. *Cell* 2008, **134**(2):317-328.
64. Ogmen U, Keskin O, Aytuna AS, Nussinov R, Gursoy A: **PRISM: protein interactions by structural matching**. *Nucleic Acids Res* 2005, **33**(Web Server issue):W331-336.
65. Tuncbag N, Keskin O, Gursoy A: **HotPoint: hot spot prediction server for protein interfaces**. *Nucleic Acids Res 2010*, **38**(Web Server issue):W402-406.
66. Darnell SJ, Page D, Mitchell JC: **An automated decision-tree approach to predicting protein interaction hot spots**. *Proteins* 2007, **68**(4):813-823.
67. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L: **The FoldX web server: an online force field**. *Nucleic Acids Res* 2005, **33**(Web Server issue):W382-388.
68. Magrane M, Consortium U: **UniProt Knowledgebase: a hub of integrated protein data**. *Database (Oxford)* 2011, **2011**:bar009.
69. Berman H, Henrick K, Nakamura H: **Announcing the worldwide Protein Data Bank**. *Nat Struct Biol* 2003, **10**(12):980.
70. Chothia C, Lesk AM: **The relation between the divergence of sequence and structure in proteins**. *Embo J* 1986, **5**(4):823-826.
71. Arnold K, Bordoli L, Kopp J, Schwede T: **The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling**. *Bioinformatics* 2006, **22**(2):195-201.
72. Pieper U, Webb BM, Barkan DT, Schneidman-Duhovny D, Schlessinger A, Braberg H, Yang Z, Meng EC, Pettersen EF, Huang CC *et al*: **ModBase, a database of annotated comparative protein structure models, and associated resources**. *Nucleic Acids Res 2008*, **39**(Database issue):D465-474.
73. Kelley LA, Sternberg MJ: **Protein structure prediction on the Web: a case study using the Phyre server**. *Nat Protoc* 2009, **4**(3):363-371.
74. Biegert A, Mayer C, Remmert M, Soding J, Lupas AN: **The MPI Bioinformatics Toolkit for protein sequence analysis**. *Nucleic Acids Res* 2006, **34**(Web Server issue):W335-339.

75. Aytuna AS, Gursoy A, Keskin O: **Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces**. *Bioinformatics* 2005, **21**(12):2850-2855.
76. Humphrey W, Dalke A, Schulten K: **VMD: visual molecular dynamics**. *J Mol Graph* 1996, **14**(1):33-38, 27-38.
77. Maemura K, de la Monte SM, Chin MT, Layne MD, Hsieh CM, Yet SF, Perrella MA, Lee ME: **CLIF, a novel cycle-like factor, regulates the circadian oscillation of plasminogen activator inhibitor-1 gene expression**. *J Biol Chem* 2000, **275**(47):36847-36851.
78. Asher G, Reinke H, Altmeyer M, Gutierrez-Arcelus M, Hottiger MO, Schibler U: **Poly(ADP-ribose) polymerase 1 participates in the phase entrainment of circadian clocks to feeding**. *Cell* 2010, **142**(6):943-953.