Identification and Evolution of protein-protein interactions in circadian gene regulation

Dissertation
Zur Erlangung des Grades
Doktor der Naturwissenschaften

Am Fachbereich Biologie
Der Johannes Gutenberg-Universität Mainz

Marcel Chris Conrady
geb. am 08.02.1994 in Offenbach am Main

Mainz, 2023

Dekan: Prof. Dr. Eckhard Thines

1. Berichterstatter: Prof. Dr. Eva Wolf
2. Berichterstatter: Prof. Dr. Helen May-Simera

Tag der mündlichen Prüfung: 03.05.2024

Zusammenfassung

Die circadiane Uhr ist ein evolutionär konservierter Zeitmessmechanismus, der es Organismen erlaubt tägliche Umweltänderungen zu antizipieren. In Säugetieren aktiviert der CLOCK:BMAL Komplex die Transkription der Repressoren CRY1/2 (CRY-m) und PER1/2, die im Cytosol einen Komplex bilden und nach Kernimport die CLOCK:BMAL Aktivität unterdrücken, resultierend in einem 24 Stunden Rhythmus. Störungen dieses Netzwerks können zu verfrühtem und verspätetem Schlafphasensyndrom führen und sind relevant für (metabolische) Erkrankungen und Krebs.

Elemente und Struktur der circadianen Uhr sind konserviert in Tieren und die ursprüngliche Struktur ist vermutlich im Monarch Schmetterling (*Danaus plexippus*) erhalten, der zusätzlich ein TIM-d und ein Licht-sensitives CRY (CRY-d) hat. In der Fruchtfliege (*Drosophila melanogaster*) ging CRY-m verloren, stattdessen reprimiert der PER:TIM-d Komplex.

In Säugetieren ist die Repression abhängig von der Interaktion von CRY-m mit der BMAL Transaktivierungsdomäne (TAD). Die Literatur zu dieser Interaktion ist jedoch widersprüchlich. Höhere Affinität und Stabilität der Repressoren führt bekanntermaßen durch verlängerte Repressionsdauer zu Phasenverzögerung. Hier identifizierten wir die BMAL-TAD Bindestelle in oder in der Nähe der primären Tasche des CRY-m durch bioinformatische, biochemische und zellbiologische Methoden. Desweitern, entdeckten wir eine Bindepräferenz für negativ geladene BMAL-TAD Aminosäuren, die konserviert ist zwischen Insekten und Säugetier CRY-m Homologen.

Zusätzlich wurde die Konkurrenz zwischen dem Ko-Repressor PER und der BMAL-TAD um CRY-m charakterisiert in Homologen aus Maus und Monarch. Konkurrierende PER Paralog-spezifische Epitope wurden u. a. im C-terminalen Bereich von PER identifiziert und durch Zellbiologie validiert. Die Konkurrenz scheint über Maus und Monarch hinaus konserviert zu sein, da die PER Proteine aller Insekten konservierte Monarch-ähnliche C-terminalen Sequenzen aufweisen und die Paralog-spezifischen Epitope von Maus PER1 und PER2 auch in anderen Wirbeltieren vorliegen. Diese Konkurrenz unterstützt das konsekutive, mehrstufige Modell der Repression, das in Säugetieren beschrieben ist und hebt die Relevanz von PER in der früheren repressiven Phase hervor. Interessanterweise ist in *Drosophila* nur ein der früheren repressiven Phase ähnlicher Repressionsmechanismus bekannt.

Die umfassende bioinformatische Analyse von Uhrnetzwerken von Insekten und Seesternen bis hin zu den Säugetieren enthüllte eine hohe Konservierung der meisten Protein-Interaktionsstellen, zeitgleichen Verlust bestimmter Interaktionsstellen und die weite Verbreitung der ursprünglichen Monarch-ähnlichen Uhr. Die Honigbienen-ähnliche Uhr, die durch einen Verlust von CRY-d und TIM-d charakterisiert ist, findet sich nur in eusozialen Insekten Ordnungen (staatenbildende Insekten). Wir fanden Hinweise für konvergente Evolution der Temperaturkompensation in Chordaten und Insekten, die auf mehreren konkurrierenden die PER Stabilität kontrollierenden Phosphorylierungsstellen basiert.

Diese Studie beschreibt ein neues Modell für die CRY-m:BMAL-TAD Interaktion, eine konservierte Konkurrenz zwischen PER and BMAL um CRY-m und funktionelle Unterschiede zwischen den C-terminalen Bereichen der Säugetier PER Paraloge, welche vermutlich zur ihrer Funktion in der Uhr beitragen. Die Evolutionsstudie der circadianen Uhr legt die konvergente Evolution der PER Stabilitätsregulation und die Korrelation von Uhrstruktur und Sozialität offen, und hebt die Konservierung der CRY-m:BMAL-TAD Interaktion und die Konkurrenz hervor.

Abstract

The circadian clock is an evolutionary conserved timekeeping mechanism that allows organisms to anticipate daily environmental changes. In mammals, CLOCK:BMAL drives transcription of CRY1/2 (CRY-m) and PER1/2 which form a complex in the cytosol and repress CLOCK:BMAL activity after nuclear import which results in a 24 h rhythm. Disruptions of this network can lead to advanced- or delayed sleep phase syndrome and are relevant in (metabolic) diseases and cancer.

Elements and structure of the circadian clock are conserved in animals and the ancestral structure is believed to be present in the monarch butterfly (*Danaus plexippus*) which additionally has a TIM-d and a light sensitive CRY (CRY-d). In the fruit fly (*Drosophila melanogaster*) the CRY-m is lost, instead the PER:TIM-d heterodimer represses.

In mammals, repression critically depends on the interaction of CRY-m with the BMAL transactivation domain (TAD). Yet, the literature on this interaction is conflicting. Higher affinity and stability of repressors are known to lead to phase delay by increasing the duration of repression. Here we identified the BMAL-TAD binding site to be in or close to the CRY-m primary pocket by using bioinformatics, biochemistry and cell biology. Furthermore, we uncovered a binding preference for negatively charged BMAL-TAD residues, that is conserved between insect and mammalian CRY-m homologues.

Additionally, the competition between the co-repressor PER and the BMAL-TAD for CRY-m was characterized in mouse and monarch homologs. Competing PER paralog-specific epitopes were identified in the C-terminus and validated by cell biology. The competition appears to be conserved beyond monarch and mouse since all insects have a conserved monarch-like C-terminal PER tail and the paralog-specific epitopes of mouse PER1 and PER2 are also present in other vertebrates. This competition supports the consecutive multi-step model of repression described in mammals and highlights the relevance of PER in the early repressive phase. Interestingly, only an early repressive phase-like mechanism is described in *Drosophila*.

The distribution of clock networks was investigated from insects and sea urchins to mammals and revealed high conservation of most interaction sites, simultaneous loss of binding sites and widespread presence of the ancestral monarch-like clock. A honeybee-like clock, characterized by loss of CRY-d and TIM-d, is only found in eusocial insect orders (e. g. colony-building insects). We found evidence for convergent evolution of temperature compensation in chordates and insects which is based on multiple phosphorylation sites that control PER stability.

This study describes a new model for the CRY-m:BMAL-TAD interaction, a conserved competition between PER and BMAL for CRY-m and functional differences between the tails of mammalian PER paralogs which are likely to contribute to their clock function. The evolutionary study of the circadian clock exposed a convergent evolution of PER stability regulation, a correlation of clock structure and sociality, and highlights the conservation of the CRY-m:BMAL-TAD interaction and the competition by PER.

Table of contents

Zusammenfassung ii
Abstract iii
Table of contents 1
List of Abbreviations 5
List of Figures 8
List of Tables 10
1 Introduction 11
1.1 Biological Clocks 11
1.1.1 Geophysical rhythms working on biology 11
1.1.2 The circadian clock phenotype 13
1.1.3 Physiological/anatomical basis of the circadian clock 13
1.1.4 Molecular basis 14
1.1.5 Pathophysiological relevance of the circadian clock 15
1.1.6 The circadian clock is essential in animal navigation 16
1.2 Interactions within the TTFL 17
1.2.1 Activation 17
1.2.2 Repression 21
1.2.3 Synchronization 29
1.2.4 TTFL models /network structures 30
1.3 Photolyase to Cryptochrome: from UV repair to UV avoidance 31
1.4 Open questions in the circadian central loop 33
1.5 Aim of the thesis 35
1.5.1 Identify the BMAL TAD binding site on CRY-m 35
1.5.2 Investigate the PER/BMAL competition for CRY-m binding under the current model 35
1.5.3 Investigate evolution of the clock network and its defining interaction sites 35
2 Materials 36
2.1 36
3 Methods 47
3.1 Tree construction 47

Table of contents

3.2 Protein structure prediction by AlphaFold multimer 47

3.3 RNA extraction 47

3.4 cDNA synthesis 48

3.5 Molecular Cloning 48

3.6 Preparing chemically competent *E. coli* 48

3.7 Transformation of chemically competent *E. coli* 49

3.8 Mini-prep of plasmid DNA from *E. coli* 49

3.9 PCR 49

3.10 Colony PCR 50

3.11 Agarose gel electrophoresis 51

3.12 Protein expression in *E. coli* 51

3.13 Blue/white screening & Bacmid DNA purification 51

3.14 Transfection into Sf9 cells and viral stock production 52

3.15 P1 expression check 52

3.16 Protein expression in *Sf9* insect cells 53

3.17 SDS-PAGE 53

3.18 Western blot 54

3.19 Protein Purification 54

3.19.1 Cell lysis 54

3.19.2 Immobilized metal ion affinity chromatography (IMAC) 55

3.19.3 GST Affinity chromatography 55

3.19.4 Ion-exchange chromatography 56

3.19.5 Size exclusion chromatography 56

3.20 DNA and protein quantification 57

3.21 Isothermal Titration Calorimetry (ITC) 58

3.22 Fluorescence Polarization 58

3.23 Crosslinking 59

3.24 Photo-Crosslinking 59

Table of contents

3.25 Photoaffinity labeling 60
3.26 Mass-spectrometry and data analysis (Collaboration) 60
3.26.1 In-gel digestion 60
3.26.2 Liquid chromatography tandem mass spectrometry 61
3.26.3 Mass spectrometry data analysis 61
4 Results 62
4.1 The CRY-m:BMAL1 TAD interaction 62
4.1.1 AlphaFold 62
4.1.2 Photoaffinity labeling 65
4.1.3 Mutants 66
4.1.4 Evolutionary conservation of interaction 68
4.2 Competitive binding of PERIOD and BMAL1 TAD to repressor CRY-m 76
4.2.1 The case of monarch butterfly CRY2:PER 76
4.2.2 PER reduces CRY-m:BMAL TAD affinity 77
4.3 Evolutionary analysis of interactions sites in the circadian clock central loop 88
4.3.1 Study design – species set 88
4.3.2 Interaction site definitions used 90
4.3.3 Circadian network structure is partially order specific 92
4.3.4 CRY types and interaction sites define network structure 93
4.3.5 Interaction sites conservation 97
4.3.6 Summary 102
5 Discussion 105
5.1 The CRY-m:BMAL1 TAD interaction 105
5.1.1 Critical literature review 105
5.1.2 Where does BMAL1 TAD bind to CRY? 107
5.1.3 Evolution of interaction 111
5.1.4 Activatior preference are more complex 111
5.2 Competitive binding of PERIOD and BMAL1 TAD to repressor CRY1 112
5.2.1 Model of competition 112

5.2.2 Why was this effect not observed earlier? 114
5.2.3 How does this change the model of the central loop? 115
5.3 Evolutionary analysis of interactions sites in the circadian clock central loop 117
5.3.1 CRY interaction sites evolution/conversation 117
5.3.2 Do we need to look for interaction sites or is gene presence a sufficient indicator? 117
5.3.3 Conservation of PER:BMAL competition 118
5.3.4 Need for multiple CRY-m and PERs 118
5.3.5 Conservation of FASP/PER short 119
5.3.6 The curios case of higher Diptera 119
5.3.7 The polyQ in Anopheles 120
5.3.8 Other networks 120
5.4 AlphaFold 121
5.5 Evolution and ecology 123
5.5.1 Chrono lifestyles and migration 123
5.5.2 Eusociality/subsociality and superorganismic needs 123
5.5.3 Chelicerata 124
5.5.4 The curios case of marine animals 124
5.5.5 Intertidal zone organism in the spotlight 125
5.6 Conclusion 126
5.6.1 The CRY-m:BMAL1 TAD interaction 126
5.6.2 Competitive binding of PERIOD and BMAL1 TAD to repressor CRY 126
5.6.3 Evolutionary analysis of interactions sites in the circadian clock central loop 127
6 Bibliography 128
Annex A
Acknowledgements L
CV N

List of Abbreviations

ARNT	-	Aryl hydrocarbon receptor nuclear translocator
ARNTL1	-	Aryl hydrocarbon receptor nuclear translocator-like protein 1
ARNTL2	-	Aryl hydrocarbon receptor nuclear translocator-like protein 2
bHLH	-	basic helix-loop-helix
BMAL1	-	Basic helix-loop-helix ARNT-like protein 1/Brain and muscle ARNT-like 1
BMAL2	-	Basic helix-loop-helix ARNT-like protein 2/Brain and muscle ARNT-like 2
BS3	-	Bis(sulfosuccinimidyl)suberate
cAMP	-	cyclic adenosine monophosphate
CatEx	-	cation exchange chromatography
CBD	-	CRY binding domain (of PER)
CBP	-	CREB binding protein
cDNA	-	complementary DNA, the DNA reverse transcribed from an RNA template
ChIP	-	Chromatin Immunoprecepitation
CIPC	-	CLOCK interacting pacemaker
CLOCK	-	CLOCK LOCOMOTER OUTPUT CIRCADIAN KAPUT protein
CK1	-	Casein kinase 1
CREB	-	cAMP responsive element binding protein
CRY	-	CRYPTOCHROME protein
CV	-	column volume
DBT	-	DOUBLE TIME protein, *Drosophla* homologue to CK1
DMSO	-	Dimethyl sulfoxide
DSPD	-	delayed sleep phase disorder
E-box	-	enhancer box
EDTA	-	Ethylenediaminetetraacetic acid
FAD	-	flavin adenine dinucleotide
FASP	-	familial advanced sleep phase
FBXL3	-	F-box/Leucine rich repeat protein 3
fl	-	full-length protein
FP	-	fluorescence polarization
FPLC	-	Fast protein liquid chromatography
HAT	-	Histone acetyltransferase

IEX	-	ion exchange chromatography
IMAC	-	immobilized metal ion affinity chromatography
IPTG	-	isopropyl-β-D-thiogalactopyranoside
iTOL	-	interactive Tree of Life, online tool to plot data on phylogenies
kDa	-	kilodalton
K_D	-	dissociation constant
LB	-	Luria Bertani Medium
MLL1	-	MIXED LINEAGE LEUKEMIA1 protein
MOPS	-	3-(N-morpholino)propanesulfonic acid
MW	-	molecular weight
MSA	-	Multiple sequence alignment
NCBI	-	National Center of Biotechnology Information
NES	-	nuclear export sequence
NHS	-	N-Hydroxysuccimide used in activation of carboxyl groups
NLS	-	Nuclear localization sequence
nr	-	non-redundant (sequence databases of the NCBI)
Nt	-	nucelotides
PAL	-	Photoaffinity Labeling
PAS	-	PER-ARNT-SIM
PCR	-	polymerase chain reaction
PER	-	PERIOD protein
PHR	-	photolyase homology region
PMSF	-	Phenylmethylsulfonyl fluoride
SCN	-	suprachiasmatic nucleus
SDS	-	sodium dodecylsulfate
SDS-PAGE	-	sodium dodecylsulfate polyacrylamide gel electrophoresis
SEC	-	Size exclusion chromatography
SPG buffer	-	Succinic acid, sodium dihydrogen Phosphate, Glycine buffer
Sulfo-SDA	-	sulfosuccinimidyl 4,4'-azipentanoate; Sulfo-NHS-Diazirine
TAD	-	Transactivation domain
TAE	-	Tris-acetate-EDTA buffer
TB	-	Terrific broth
TCEP	-	Tris(2-carboxyethyl)phosphin

TEMED	-	Tetramethylethylenediamine
TIM	-	TIMELESS
TIPIN	-	TIMELESS-interacting protein
TRX	-	TRITHORAX protein, *Drosophia* and monarch homologue to MLL1
TSA	-	Transcriptome Shotgun Assembly
TSA	-	Thermal Shift Assay
TTFL	-	transcription translation feedback loop
WBM	-	WDR5 binding motif
WIN	-	WDR5 interacting
WT	-	wild-type
X-Gal	-	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
XL-MS	-	Crosslinking-mass spectrometry
YFP	-	yellow fluorescent protein

List of Figures

Figure 1: A negative feedback loop of 24 h periodicity and influencing factors. 11
Figure 2: Biological rhythms imposed by celestial bodies. 12
Figure 3: Example double plotted actogram 13
Figure 4: Simplified clock networks 15
Figure 5: Circadian rhythms in bees (left) and monarch migration (right). 17
Figure 6: Interactions of the BMAL1:CLOCK heterodimer. 20
Figure 7: Repressive interactions. 23
Figure 8: Differences in CRY-m and CRY-d structures and conservation. 26
Figure 9: Repressive interactions in the Drosophila clock. 29
Figure 10: Mammalian and Drosophila clocks are derivatives of the ancestral clock 30
Figure 11: Evolution of the CRY/PL family. 34
Figure 12: Sulfo-SDA reaction scheme. from product page 59
Figure 13: Scheme of modified peptide. Image courtesy by Dr. Fabian Barthels. 60
Figure 14: Results of CRY1:BMAL1 AlphaFold multimer prediction and XL-MS. 64
Figure 15: Results of the CRY1 mutants. 69
Figure 16: Result of BMAL TAD variant and insect CRY-m interaction. 72
Figure 17: Alignment of the CRY-m used. 75
Figure 18: Monarch butterfly PER reduces CRY:BMAL TAD interaction 78
Figure 19: Competition of mouse PER1/2 with BMAL1 TAD for CRY1/2 binding. 80
Figure 20: CRY1:PER1/2 crosslinks 84
Figure 21: CRY2:PER1/2 crosslinks 85
Figure 22: Basis of competition by PER1/2 87
Figure 23: Protein homologs and clock networks. 89
Figure 24: Interaction sites and clock network structures are conserved 94
Figure 25: Multiple sequence alignment of the PER CBD and C-terminus. 99
Figure 26: Conservation of per^{s} site. 102
Figure 27: Lengths of CRY-m and PER tails. 104
Figure 28: Simplistic model for CRY CC:BMAL TAD interactions. 106
Figure 29: BMAL co-activator preferences. 112
Figure 30: Model of PER1 tail (orange) repression. 113
Figure 31: Alignment of PER CBDs 114
Figure 32: Updated model. 116
Figure 33: Alignment of Clunio BMAL TAD isoforms with TADs of selected animals. 126

List of Figures

Figure 34: Purification of insect CRY-m PHR. ..... D
Figure 35: MSA CRY-m, CRY-d. ..... E
Figure 36: MSA CRY-m, CRY-d. ..... F
Figure 37: Multiple sequence alignment of mouse and insect CRY-m used in this thesis. ..... G
Figure 38: ITC result of CRY:PER interaction. ..... H
Figure 39: AlphaFold multimer predictions with prediction scores. I
Figure 40: Sequence logo of PER and CLOCK motifs ..... J
Figure 41: The FASP region for insect orders, Chordata and marine organisms. ..... K

List of Tables

Table 1: Differences of CRY paralogs. 25
Table 2: Differences of PER paralogs. 28
Table 3:Buffers 36
Table 4: Media 36
Table 5:Purification buffers 37
Table 6: Equipment 37
Table 7: Special Chemicals and Consumables 39
Table 8: Western blot 39
Table 9: RNA work 39
Table 10: Columns and affinity resins 40
Table 11: Commercial Kits 40
Table 12: Standards 40
Table 13: Peptides 41
Table 14:Plasmid backbones 41
Table 15: Software 42
Table 16: Webtools 42
Table 17: Primers 42
Table 18: Enzymes 45
Table 19: CRY Proteins 45
Table 20: PER proteins 46
Table 21: Cells 46
Table 22: cDNA synthesis 48
Table 23: Pipetting scheme of high-fidelity PCR. 49
Table 24: Colony PCR pipetting scheme 50
Table 25: Pipetting scheme for SDS-PAGE gel preparation 53
Table 26: Purification steps for all proteins used 57
Table 27: Results of the mutant study. 68
Table 28: PER CBD and CBD+tail constructs used 82
Table 29: Investigated interaction sites and the reference sequences used to identify them. 92
Table 30: Species with a CRY-m but no PER CBD detected 95
Table 31: Species with CRY-m but no BMAL TAD 96
Table 32: Phospho-sites in the FASP region by insect order. 101
Table 33: Composition of clock networks 117

1 Introduction

1.1 Biological Clocks

Circadian clocks are molecular oscillators with ~24 h periodicity (circa = about, dies = day) which allow organisms to anticipate daily environmental changes such as day/night and associated changes in temperature. The general structure consists of a negative feedback loop in which an activator controls the activity of its repressor (Figure 1) and depending on dissociation constants, protein concentrations, post translational modifications (PTM), transcription, translation, translocation and degradation kinetics a continuous loop of constant period and amplitude establishes. Circadian clocks are found in all clades of life[1–3].

Figure 1: A negative feedback loop of 24 h periodicity and influencing factors.

1.1.1 Geophysical rhythms working on biology

External rhythms are imposed on biological systems on our planet primarily by two celestial bodies, the moon and the sun. Examples of geophysical rhythms are shown in Figure 2 with accompanying biological rhythms.

1.1.1.1 Circadian rhythms

The most obvious and one of the strongest geophysical rhythms impacting biological systems is the 24 h day-night cycle resulting from earth's rotation around its axis. The day-night cycle dictates activity patterns in animals and phases of photosynthetic activity in plants and cyanobacteria. The day night cycle also brings along an increase in UV radiation during the day which was even more important in the early ozone and oxygen lacking atmosphere. Temperature is dependent on sun position. Activity patterns of most animals in turn are also dependent on temperature. For example, most animals are poikilotherm meaning they don't generate their own body heat and depend on environmental temperatures for activity.

1.1.1.2 Circalunar rhythms

Many marine organisms are known to use the moon phase to synchronize reproductive efforts. Reproduction is energy costly, especially egg cell production. Sessile organisms like corals cannot physically meet to exchange genetic material at high concentrations increasing the chances of successful fertilization. Interestingly, not only sessile organisms eject egg and sperm for external fertilization but also motile ones like the annelid (marine worm) *Platynereis dumerilii*[4–6]. Gamete production and release are carefully timed to increase fertilization probability by increased

concentrations of egg and sperm by synchronous release[7]. This strategy is so important that coral species diversification can potentially be driven by advanced or delayed gamete release [8].

1.1.1.3 The intertidal zone

Some geophysical rhythms only effect animals in a certain biological niche. The intertidal zone is one such example. About every 12 h the habitat is flooded and 12 h later the same area is again (relatively) dry land. This environment is conceivable challenging but specialized animals thrive in this environment such as the midge (fly) *Clunio marinus*. *Clunio* adults eclose every 15 days at spring tide. The males fertilize the females who lays its eggs in wet sand that will be covered by water during a usual ebb tide. Pupa development takes three to four days and adult midges only live a few hours. Therefore, pupation and eclosion have to happen at the right day (3-4 days prior to spring tide) and time (when water is almost fully retracted), respectively. Further, depending on geographic location, tides come at different times. Crosses between different strains show genetic determination of circadian and circasemilunar rhythms[4,9].

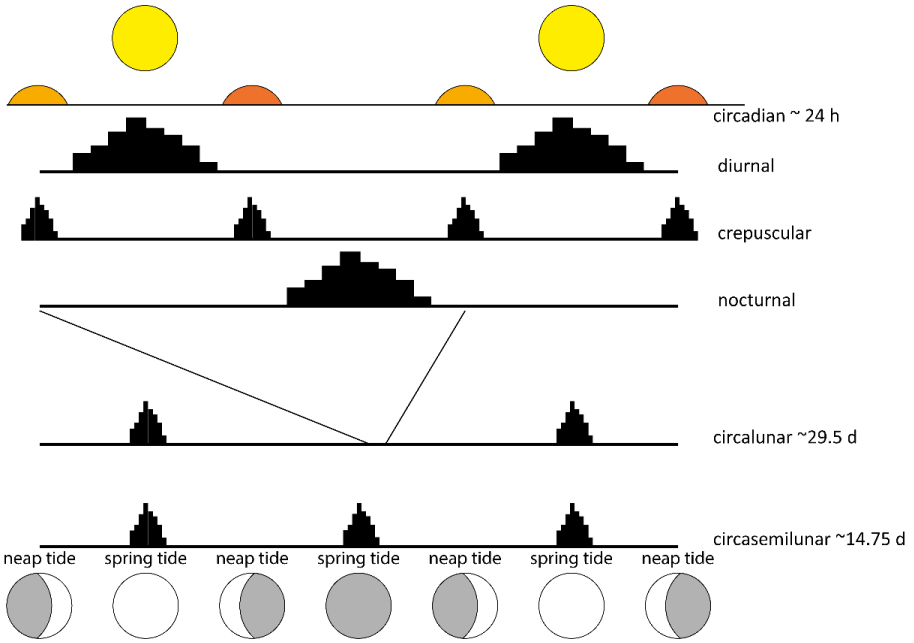

Figure 2: Biological rhythms imposed by celestial bodies. The different types of biological rhythms imposed are shown on the right. The circadian clock can result in multiple behaviors. Different clocks (e.g. circasemilunar and circadian) can work together as described above for Clunio marinus.

1.1.2 The circadian clock phenotype

Kommentiert [CM1]: Edit Bae (2001) to citation key, shorten if possible

Jürgen Aschoff is thought of as one of the fathers of chronobiology. After placing subjects in a "bunker" devoid of environmental cues like sun light and shifting temperature, Aschoff realized that humans still showed stable activity patterns of about 24 h length (circa-dian)[10]. Sun light was not a necessary cue or "Zeitgeber" to use the official terminus. Autonomy of the rhythm is the first hallmark of circadian rhythms. Secondly, circadian clocks react to changing environmental conditions, especially temperature and light allowing synchronization. The third hallmark is temperature compensation[11]. Chemical reactions and thereby most (molecular) biology follow the simple rule, the warmer the faster. Importantly, circadian clocks have to keep 24 h periodicity as the day-length doesn't change with temperature.

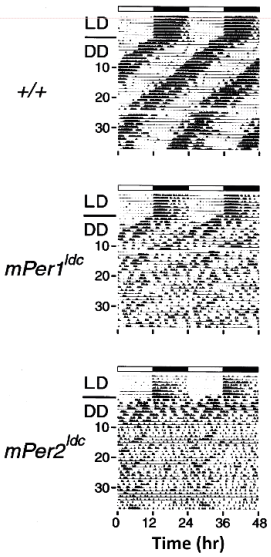

Figure 3: Example double plotted actogram of WT and PER1 and PER2 knock out mice reprinted with permission from Bae (2001)[12].

Light is an important Zeitgeber for many model organisms and can be easily manipulated. Many experiments are based on using light regimes to entrain animals. LD means light:dark and usually signifies 12 h light and 12 h darkness if nothing else is specified, while DD signifies dark:dark. A typical experiment starts with an entrainment period in LD and continues with DD to measure the free running period similar to the bunker experiment above. Activity is plotted against time and light regime (Figure 3; from [12]). Activity is measured in mice by wheel-running activity. The monarch butterfly is also an important model organism in chronobiology partly because of its seasonal migration pattern. Monarchs are glued to a lever and flight/activity is measured by the pull they exert. If seasonality is under investigation the LD regime is typically modified to recapitulate summer or fall/winter with light duration depending on the relevant latitude (e.g. summer 16:8, winter 8:16). Directional flight induced by seasonal changes is measured in the monarch butterfly by recording the direction in which the lever is pulled. In LD experiments time is often given as Zeitgeber time, ZT, where lights on defines ZT 0 and onset of activity for diurnal animals. ZT 12 (lights off) defines onset of activity in nocturnal animals. Circadian time on the other hand is based on the free running period (in DD). CT 0 and CT 12 is defined as onset of activity in diurnal and nocturnal animals, respectively.

1.1.3 Physiological/anatomical basis of the circadian clock

Later, it was discovered that the circadian clock is not only present on the organism level but also on the cellular level. Cultured cells will usually be asynchronous but can be synchronized [13–15]. The circadian clock network is therefore not based on neural circuits as our perception of time[16] but instead

is cell autonomous.. To keep all tissues and organs synchronized a central pacemaker is needed. The central pacemaker is located in the suprachiasmatic nucleus (SCN). The SCN sits at the crossing of the optic nerves allowing synchronization to the Zeitgeber light perceived in special photoreceptors in the eye and transmitted by the optic nerves[17–23]. The pineal gland produces melatonin and in some species of lampreys, fishes, amphibians, reptiles and birds retains photosensitivity[24]. The SCN clock neurons synchronize with each other generating a very robust central pacemaker[25]. Peripheral body clocks such as in the liver are synchronized by direct efferent projections or hormones[26]. Other strong Zeitgebers that can reset peripheral clocks are calory intake[27] and physical activity[28,29]. In the monarch butterfly the central pacemaker in the brain is located in only four cells in the pars lateralis and additional circadian clocks reside in the antenna[30].

1.1.4 Molecular basis

The molecular basis of this cell-autonomous, temperature compensated, and robust pacemaker is a transcription-translation feedback loop (TTFL). This thesis is focused on animal clocks and so will this chapter. Precisely, the intensively studied mammalian and *Drosophila* systems will be explained. Plants[31] and fungi[32] also use one or multiple interlocking TTFLs. In cyanobacteria the clock is based on two to three proteins (KaiAB(C)) and a phosphorylation cycle[3]. A great tool to study because it can be entirely reconstituted and studied in a reaction tube for weeks[33,34].

The TTFL is based on a negative feedback loop. In animals the activator consists of a heterodimer of CLOCK and BMAL. CLOCK (Circadian locomotor output cycles protein kaput) and BMAL (Brain and Muscle Arntl-like) together bind circadian promoters at enhancer boxes (E-Box)[35,36]. Around CT4-8 CLOCK:BMAL recruit co-activators whose enzymatic activity results in transcriptionally active, open chromatin and recruitment of transcription machinery[37]. CLOCK:BMAL activity leads to expression of many clock controlled genes (ccg) and importantly also of PERIOD (*per*) and CRYPTOCHROME (*cry*). PER and CRY proteins form a complex in the cytosol and translocate into the nucleus together with casein kinase 1 (CK1). They accumulate and bind to CLOCK:BMAL repressing their activity around CT12-20 in what is called the early repressive phase. Not all components of this large complex are known[37,38]. PER-recruited CK1 phosphorylates not only PER but also PER associated proteins[39]. In the late repressive phase the repressor complex disassembles and CRY1 binds to CLOCK:BMAL keeping them in a poised state[37] until CT4. Phosphorylation of PER leads to its degradation[40] and due to the loss of PER protection CRY is degraded as well[41,42]. In the absence of CRY and PER, CLOCK and BMAL will bind E-boxes again and activate transcription. CT values refer to the situation in the mouse liver.

Drosophila lacks a full length BMAL protein. A shorter BMAL homolog, called CYCLE is present instead. Transactivation by CLOCK:CYCLE is much less understood. *Drosophila* misses a repressor CRY[43]. CLOCK:CYCLE repression is based on a PER:TIM-d heterodimer that recruits the CK1 homolog

DOUBLETIME (DBT). *Drosophila* contains a light sensitive CRY (photoreceptor CRY). Upon blue light irradiation CRY binds TIM, freeing PER, resulting in degradation of these proteins via recruitment of the E3 ubiquitin ligase JETLAG[44–46] and SLIMB[47–49], respectively. The *clock* gene is expressed rhythmically in *Drosophila*, while *bmal1* is expressed rhythmically in mammals[50]. Importantly, the mammalian and *Drosophila* system share components and mechanisms but activation and repression work differently. Further, multistep repression is not described yet for insect clocks.

Another molecular mechanism exists which was identified later in the monarch butterfly. This mechanism is essentially a mix of the two aforementioned because it is actually the ancestral system[43].

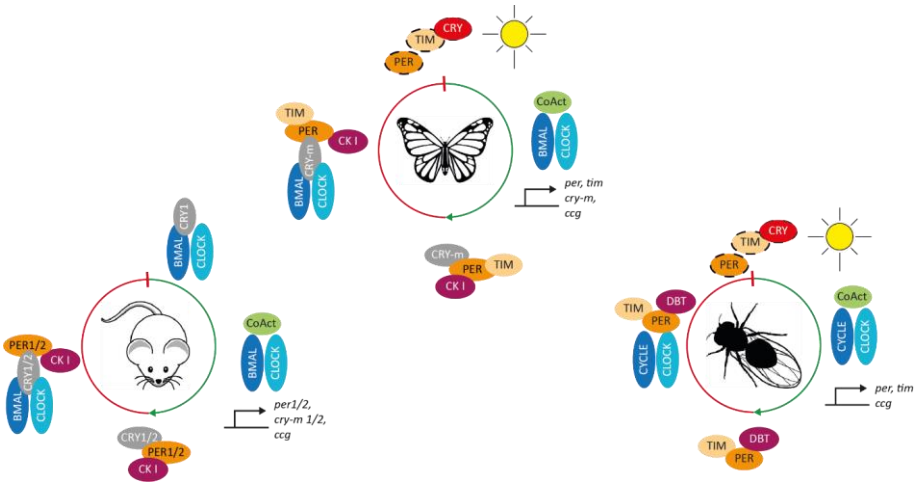

Figure 4: Simplified clock networks of mammals (left), Drosophila (right) and the ancestral monarch clock (top). Ccg: clock controlled genes. Colors match for homologous proteins. Protein complex position correlates with CT as far as possible but might vary with tissue. Drosophila-like photoreceptor CRY is shown in red. Mammal-like transcription repressor CRY in grey.

Drosophila-like photoreceptor CRYs and mammal-like transcription repressor CRYs will be called CRY-d and CRY-m from hereon. The data presented here stems from studies on mouse proteins/genes if not otherwise stated. Mouse and human clock proteins are almost identical on sequence level and molecular biology results are very transferable. Behavioral studies are different as humans are diurnal and mice nocturnal.

1.1.5 Pathophysiological relevance of the circadian clock

Depending on the source 25% up to the majority of mammalian genes exhibit daily rhythms in transcription levels making the circadian clock one of the most important gene regulatory networks. Therefore, pathophysiological processes are influenced or caused by the circadian clock like sleep

disorders, cancer, and metabolic diseases. Expression of many metabolic enzymes is under circadian control[51]. On the other hand, timing of drug administration for the same reason can have a huge impact on therapy success and side effects. Additionally, our modern lifestyles desynchronize us from our internal clocks by shift work, irregular sleep or frequent travelling across time zones, resulting in jetlag. Further, mobile device usage exposing us to high blue light intensities late in the evening/night, unhealthy and untimely food consumption lead to stress and metabolic diseases such as diabetes[52–54]. For example, CRY-m double knockout mice are not only arrhythmic but also show symptoms of hyperglycemia[55,56].

I want to describe two disorders which are directly caused by mutations in the central feedback loop: familial advanced sleep phase (FASP) syndrome and delayed sleep phase disorder (DSPD)[57–62]. FASP is caused by a S662G mutation in PER2. This serine starts a chain reaction of PER2 stabilizing phosphorylation events. In their absence, PER2 is destabilized resulting in shorter periods/advanced phases[58–61]. DSPD is caused by a dominant mutation in CRY1 leading to deletion of exon 11 (CRY1 tail) resulting in enhanced binding to CLOCK, longer repression and therefore longer periods /delayed phase[57,62].

1.1.6 The circadian clock is essential in animal navigation

Circadian clocks are essential for animals that orient by a sun compass sense. Since the suns position changes over the course of the day its position has to be put into relation to daytime (time-compensated sun compass).

1.1.6.1 Bees foraging and homing

The honeybee, *Apis mellifera*, collects honey from a wide area around the hive (<1km up to 8km). Bees use their time-compensated sun compass to find the best nectar sources and the way back to the hive. Bees can sense the polarization of UV light allowing them to see the sun positions even in cloudy conditions. Bees show remarkably plastic rhythms: foragers show strong rhythms while brood-caring nurses have dampened amplitudes and phenotypes as shown in Figure 5 (left). Colony requirements supersede what molecular biology dictates in the superorganism of the eusocial colony[63,64].

1.1.6.2 Monarch migration

Monarch butterflies extend their range from spring to fall from central Mexico to the US/Canadian border following the growth of milkweeds on which their larvae feed. After a cold cue and decreasing day length in fall, the adult butterflies migrate in giant swarms to Mexico where they overwinter and go into diapause to start the first generation once it warms up again in spring (Figure 5, right). The long-distance migration is enabled by a time compensated sun compass. Sun positional information from the eyes is combined with time information from the antenna clock resulting in directional flight[65]. The

central pacemaker in the brain is not used for time compensation[66]. Precision of the circadian clock is of uttermost importance for correct direction over such long distances.

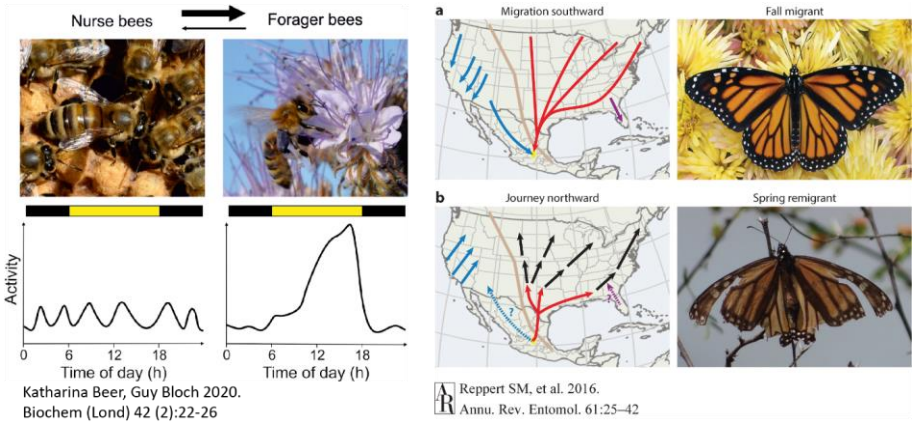

Figure 5: Circadian rhythms in bees (left) and monarch migration (right). Reprinted with permission from Beer and Bloch (2020) and Reppert et al (2016).

1.2 Interactions within the TTFL

This thesis will revolve around the identification of interactions sites, their competition and evolution of such sites within the gene regulatory network of the circadian clock. The most important and within this thesis studied interaction sites are described in the following section and organized from activation over repressive branch to the interactions that lead to synchronization. This section concludes with the identified white spots on the interaction map.

1.2.1 Activation

1.2.1.1 BMAL interactions

1.2.1.1.1 Heterodimer formation with CLOCK

BMAL and CLOCK heterodimerize with their basic-helix-loop-helix (bHLH) and two PER-ARNT-SIM (PAS) domains (PAS-A, PAS-B, Figure 6 A, B). Heterodimer formation is necessary for enhancer box (E-Box, 5'-CACGTG-3') recognition and binding in the genome. CLOCK binds 5`-CAC-3’ and BMAL1 5’-GTGA-3’[35].

1.2.1.1.2 CBP/p300 recruitment by BMAL

The C-terminal BMAL transactivation domain (TAD, Figure 6 A) is the major activation domain as shown by bioluminescence oscillation assays in mammalian cells and random mutagenesis studies[67–69]. Interaction with the co-activator CBP/p300, a histone acetyl transferase, was identified by a yeast two hybrid assay and co-activation shown by dose-dependent enhanced reporter gene activity by a CBP

fragment corresponding to the KIX domain region[69] (Figure 6 A). This was supported by CoIP of p300 by CLOCK (as part of the CLOCK:BMAL complex) and rhythmic histone 3 acetylation at the *per1* and *per2* promoters[70]. Importance of the BMAL1 TAD for transactivation and circadian rhythmicity was demonstrated by domain swapping with corresponding domains from the paralog BMAL2. Association of the BMAL1 TAD with the p300 KIX domain was shown by NMR and ITC[71]. The BMAL1 TAD does not assume a tertiary structure and the KIX domain has a molten globule fold hindering structural analysis[72]. Recently also the CBP/p300 TAZ domains were described as BMAL1 TAD binding sites (GRC Chronobiology 2023, talk by Dr. Priya Crosby). The TAZ domains are known transactivation domain binding sites[73]. CBP/p300 are well known transcriptional activators, they acetylate histones, regulate transcription via chromatin remodeling, and acetylate and recruit transcriptional machinery[74,75].

1.2.1.2 CLOCK interactions

1.2.1.2.1 Exon19

In mice, a single knockout (KO) of mouse CLOCK exon 19 leads to slightly longer period, double KO leads to substantially longer (1-4 h) period with decreased amplitude or arrhythmicity[76,77]. Interaction of CLOCK exon 19 was demonstrated by CoIP with the histone methyltransferase MLL1 complex. The exact mechanism is still unclear. Endogenous MLL1 is enriched on CLOCK:BMAL only around CT24 as shown by CoIP with CLOCK while both are constitutively expressed[76]. MLL1 primarily trimethylates H3K4 which is associated with transcription activation. In mice, a MLL1 knockout leads to arrhythmicity[76]. The interaction was also studied in the monarch butterfly on the MLL1 homolog TRITHORAX (TRX). While a deletion of exon 19 (exon 12 in monarch CLOCK) leads to reduced transcription activation, PER is unable to repress CLOCK Δ19 activity. TRX knockdown had a similar effect as exon 19 deletion and TRX catalytic activity was necessary. From this the authors concluded that the interaction between exon 19 and MLL1 is conserved in the monarch butterfly (exon12 and TRX). PER was proposed to bind over a methylated HSP68 (in mammals HSP70), found within the complex, to the CLOCK:TRX complex leading to repression[78].

An interaction of CLOCK exon 19 with CLOCK interacting pacemaker (CIPC) is understood even on structural level. Two anti-parallel exon 19 sequences together with CIPC form a coiled-coil complex (Figure 7 A, B). This structure implies that CIPC can only repress activity of two CLOCK proteins binding to spatially close E-boxes, either tandem E-boxes or E-boxes that are close due to chromatin organization[79]. CIPC was originally described as an mammalian inhibitor without invertebrate homologs[80]. A later study challenged this view by showing invertebrate homologs but failed to convincingly show an interaction. Further, the presented homologs for CIPC and exon19 were not species matched; for some species only a CIPC but no CLOCK exon 19 was presented and vice versa[79].

1.2.1.2.2 CLOCK Q-rich and polyQ

The CLOCK C-terminus is disordered containing a glutamine (Q) rich region with very limited sequence homology to NCoA3 and yeast Esa1. NCoA3 is a member of the SRC family of histone acetyltransferases (HAT) and has a similar domain architecture as CLOCK (bHLH, PAS-A, PAS-B). NCoA3 was shown to have intrinsic HAT activity which was also shown for CLOCK. This activity requires acetyl-CoA but surprisingly no folded domain. The activity is reduced by ~50% in two independent triple mutants and a deletion mutant of the low homology region. CLOCK HAT activity is essential for regulation of PER1, potentially through regulation of H3K9 and H3K14 acetylation[81,82]. Reportedly, CLOCK acetylates BMAL at K537 which is critical for CRY1 binding[83].

It's not clear how transactivation is achieved by the *Drosophila* CLOCK:CYCLE complex. The BMAL TAD was shown in mouse and monarch butterfly to be a very strong transactivator[71,84]. The lack of such a TAD in CYCLE raises the question how CLOCK:CYCLE activates transcription. CLOCK of course recruits co-activators by itself as described above for CLOCK exon 19:MLL1. The CLOCK intrinsic HAT activity was never demonstrated for *Drosophila* CLOCK and is highly unlikely to occur in *Drosophila* CLOCK based on the alignment in the original publication demonstrating the activity [81].

The *Drosophila* CLOCK has a C-terminal poly glutamine (polyQ) motif. α-helical polyQ motifs are known to induce transcription and are also implicated in neurodegenerative diseases. This polyQ motif is discussed to compensate for the absence of a transactivation domain in CYCLE[85–87]. polyQ motifs tend to expand and contract during DNA repair. polyQ motifs interact with each other[88,89]. Yet, no interactor of CLOCK polyQ has been described to this day. By homology, the only known transactivating activity of *Drosophila* CLOCK might be by TRX recruitment.

1.2.1.3 Heterodimer:heterodimer interaction on tandem E-boxes

E-boxes are prevalent throughout the genome. CLOCK:BMAL activation was already discussed longer to work primarily on tandem E-boxes to increase specificity[90]. Tandem E-boxes are observed in circadian promoters and *in vivo* the BMAL:CLOCK-dependent genes with tandem E-boxes are the transcriptionally most active genes [91]. The recent structural evidence supports this idea that shows the interaction of two CLOCK:BMAL1 heterodimers on nucleosomes by cryo electron microscopy. The chains in the second heterodimer couldn't be assigned to BMAL or CLOCK and are therefore not modeled. The interaction of CLOCK with the nucleosome is also very visible as shown in Figure 6 [36]. This also fits well to the necessity of two CLOCK molecules for CIPC repression[79] and multiple BMAL TAD interaction domains on CBP/p300 that can bind on the same promoter (GRC Chronobiology 2023, talk by Dr. Priya Crosby).

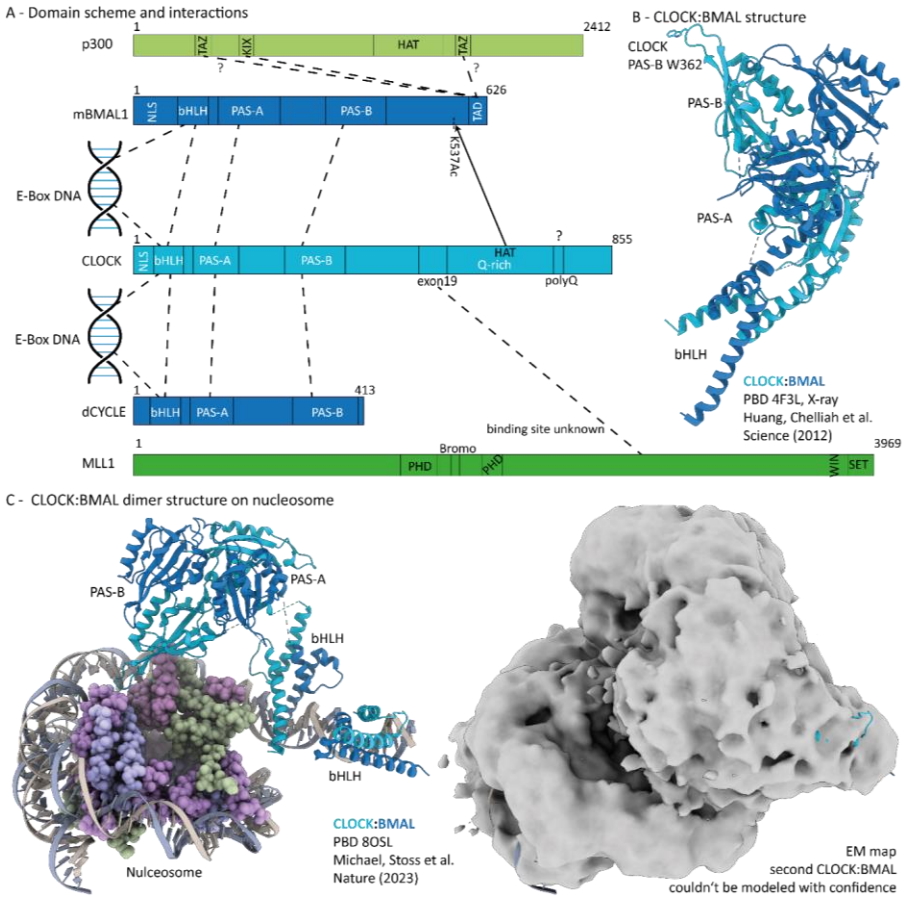

Figure 6: Interactions of the BMAL1:CLOCK heterodimer. A: Domains schemes of interacting proteins are shown. (Putative) Interactions are shown by dashed lines, enzymatic activities by arrows. Domain annotations are based on the crystal structure and Uniprot. The lengths of p300 and MLL1 domain schemes are scaled a third compared to the others. B: Crystal structure of the CLOCK:BMAL1 dimer was determined by Huang, Chelliah et al (2012)[35]. C: The structure of CLOCK:BMAL1 binding to nucleosomes was solved by Michael, Stoss et al (2023)[36] using cryo electron microscopy. CLOCK:BMAL interact with histones and two CLOCK:BMAL1 dimers interact with each other at tandem E-boxes. The second CLOCK:BMAL1 dimer couldn't be modeled with sufficient confidence, therefore the EM map is shown to highlight unmodeled density in the front, only the bHLH domain dimer is shown.

1.2.2 Repression

1.2.2.1 Repressor CRY-m interactions

1.2.2.1.1 CRY-m:CLOCK PAS-B HI W362

CRY interacts with CLOCK via tryptophan 362 located on the CLOCK PAS-B HI-loop which was shown to be important for association. The interaction can partially compensate for the loss of the BMAL1 TAD in CRY1-mediated repression[71]. The CLOCK W362A mutation leads to loss of CRY1 pulldown and co-elution in analytical SEC (as well for CLK H360Y, CRY1 R109Q, CRY1 P39A)[92]. This tryptophan sticks into the solvent in the BMAL1:CLOCK heterodimer[35] (Figure 6 B). CRY-m seems to have co-opted a conserved interaction site, as PAS:PAS dimerization in mouse and *Drosophila* PER proteins is based on the HI-loop tryptophan inserting into the partner PAS domain[35,93–95] (Figure 9). The CLOCK PAS-B HI loop tryptophan was modeled to bind into the secondary pocket of CRY-m and supported by mutagenesis studies [67,92,96] (Figure 7 C).

1.2.2.1.2 CRY-m:BMAL TAD

The BMAL transactivation domain (TAD) is required for CBP/p300 recruitment as described above. Interestingly, CRY-m binds to the same binding site leading to direct competition for binding between the co-activator and repressor. The last helix of the CRY1 PHR (CC, α22) was shown to be important for PER2 interaction, nuclear localization and efficient CLOCK:BMAL repression concluding that the CC-helix is important for repression[97]. Binding of the BMAL1 C-terminal domain to CRY1 epitopes was investigated in a peptide scan assay. Many overlapping 10 aa peptides representing the CC-helix and tail of CRY1 were tested for interaction with $BMAL1_{477-626}$. Interaction was reported on peptides representing CC-helix and tail[83].This further strengthened the idea of CRY-m CC-helix:BMAL TAD interaction (Figure 7 D). Importance of the BMAL1 TAD for transactivation and interaction with the CRY1 CC-helix were demonstrated by Xu and Gustafson et al (2015)[71]. Based on the result above the authors went straight from full length CRY to using the CRY1 CC-helix peptide to measure binding-induced chemical shifts by NMR on the BMAL1 TAD. Both CBP/p300 KIX domain and CRY1 CC-helix induced chemical shifts at the α-helical motif within the BMAL1 TAD and at the seven C-terminal residues. The conformation of mouse BMAL1 TAD bound to CRY1 CC-helix and CBP/p300 KIX domain was shown to differ. CRY1 CC-helix induces spatial proximity of the C-terminus with an attached quencher (EDTA:Mn^{2+}) to the α-helical binding epitope of the BMAL1 TAD contrary to the CBP/p300 KIX domain. Further, the affinity of the CRY1 CC-helix could be modulated by mutations in the BMAL1 TAD α-helical binding epitope. Affinity increasing mutations resulted in longer period (~1.5 h, 2-fold increase in affinity) and affinity decreasing mutations in period shortening (up to 3 h/4-fold decrease in binding). Importantly, the effect on p300 KIX binding affinity was very mild demonstrating the importance of the CRY1:BMAL1 TAD interaction for correct timing[71].

1.2.2.1.3 CRY-m:PER CBD

PER binds to repressor CRY over a C-terminal CRY binding domain (CBD). Crystal structures of this complex were solved for mouse CRY1:PER2 (PDB 4CT0)[98], CRY2:PER2 (PDB 4U8H)[99] and mouse CRY1: human PER2 (PDB 6OF7)[96]. PER wraps around the CRY-m photolyase homology region (PHR) from the secondary pocket (binding MTHF in photolyases (PL[100]) and CRY DASH cryptochromes) around the C-terminal predicted coiled-coil helix (α22) of the PHR and between the primary pocket (binding FAD in photoreceptor CRYs and PLs[100]) and helix α22 [96,98,99] burying a large solvent-accessible surface area (> 3200 Å^2, Figure 7 F,G).

The complex breaks[98,99] an intrachain disulfide bridge found in apo CRY1 (PDB 4K0R) between C363 and C412 (mouse CRY1 enumeration, Figure 7 E). This disulfide bridge is stable even in presence of reducing agents[101]. CRY-m and PER together form a Zn^{2+} binding site where mouse CRY1 provides a CH motif (C414, H473) and mouse PER2 a CxxC motif (C1210, C1213)[98,99] (Figure 7 H). Zn^{2+} binding and disulfide bridge formation are mutually exclusive due to the spatial closeness of C412 (disulfide) and C414 (Zn^{2+})[98]. Binding of Zn^{2+} stabilizes the complex and potentially orders an otherwise flexible region[99].

The redox stable disulfide bridge and mutual exclusive character suggests a connection to redox metabolism. Circadian amplitude is influenced by the cells redox state and circadian clocks influence redox systems[102].

1.2.2.1.4 PER2 binding and CLOCK

Interactions on CRY-m witch CLOCK are not independent of PER. Instead PER2 increases the binding affinity of the CRY2 paralog to the CLOCK BMAL1 PAS domain core 2-fold while decreasing the affinity of the CRY1 paralog 3-fold, bringing them in a similar range[96] (see Table 1 for K_D values). This is further supported by increased co-immunoprecipitation efficiency of CLOCK:BMAL1 by CRY2 in presence of PER2[103].

1.2.2.1.5 CRY-m:FBXL3/21

Timely degradation of CRY-m was demonstrated to be important for circadian rhythmicity.[41,104] FBXL3 and FBXL21 were identified to be E3 ubiquitin ligases targeting CRY-m[105–107]. The structure of the FBXL3 complex binding to mouse CRY2 paralog was solved (PDB 4I6J[42], Figure 7 I, J) revealing that the FBXL3 C-terminus inserts into the primary/FAD binding pocket of CRY2. FBXL3 covers a large surface area that is overlapping with the PER2 binding site, explaining why PER2 protects CRY1/2 from degradation and why PER2 bound CRY1/2 is devoid of FBXL3[42,98]. PER2 is able to release FBXL3-bound CRY2, just like FAD[42], showing again the importance of overlapping, competing binding sites in the circadian gene regulatory network.

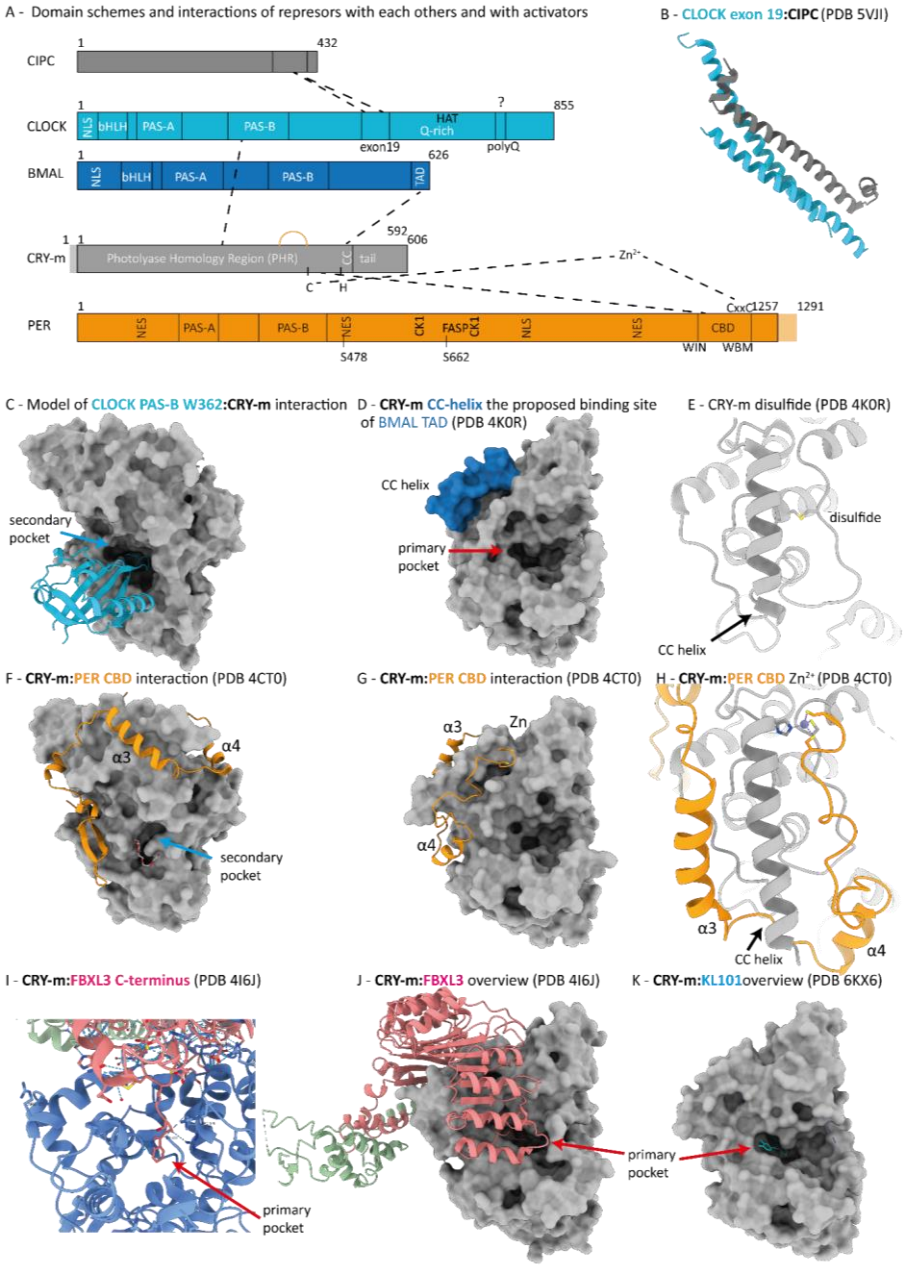

Figure 7: Repressive interactions. A: domain schemes of repressors, CRY-m and PER and their targets BMAL and CLOCK. Dashed lines indicate interactions, yellow arc on CRY-m shows the disulfide bridge. The lighter shade depicts the relative size of the mammalian paralogs CRY2 and PER1. B: Depicted is the trimeric interaction of two CLOCK exon 19 domains with CIPC. C: The proposed model for CLOCK:CRY-

m interaction is shown (AlphaFold prediction). The interaction of PER2 (orange) with CRY1 (grey) is shown in the same orientation in F. D: The binding site of the BMAL1 TAD on CRY-m is described to be on the CC-helix highlighted in blue. G and J show binding of PER2 (orange) and FBXL3 (light coral) in the same orientation, respectively. The PER2 CBD (orange) wraps tightly around the CRY1 CC-helix as shown in F-H. A comparison of disulfide bond and Zn^{2+} binding site in apo CRY1 and CRY1:PER2 complex is shown in E and H, respectively. Interactions in the CRY-m primary pocket are shown in I-K. I and J show the interaction of FBXL3 (light coral), whose C-terminus is inserting deep into the CRY2 primary pocket. J shows the CRY1 specific inhibitor KL101 binding to the CRY1 primary pocket. Structures for the homodimerization of mouse PER1-3 over PAS domains are available but are not presented here.

1.2.2.1.6 CRY-m Inhibitors

KL001 was the first identified CRY-m inhibitor. Chemically unrelated, paralog specific inhibitors of CRY1 and CRY2 exist, KL101 and TH301, respectively[108–111]. These inhibitors bind similar as KL001 to the FAD binding pocket as shown in Figure 7 and inhibit the ubiquitin dependent degradation[111] due to steric clash with the FBXL3 C-terminus required for CRY binding[42,112]. KL101 and TH301 lengthen period and dampen amplitude concentration-dependently[108].

1.2.2.1.7 CRY-m tail

CRY-m and CRY-d have a conserved PHR which is followed by a predicted disordered and divergent tail of varying length. The tails interact with the PHR as shown by tail exon 10 defining compound specificity of CRY-m inhibitors[108] and exon 11 auto-inhibiting CLOCK binding[62]. The tail was also shown to interact with the BMAL1 TAD[83]. A lack of autoinhibition in CRY1 leads to delayed sleep phase disorder (DSPD) and is caused by a deletion of CRY1 exon 11 leading to increased interaction with CLOCK:BMAL due to faster association and increased CRY1-mediated repression of CLOCK:BMAL activity[62].

1.2.2.1.8 Differences between CRY1 and CRY2

CRY2 occupancy at DNA regulatory elements increases from early circadian day, peaking at early circadian night (CT 16). This coincides with PER-mediated transcriptional repression. CRY1 binding starts and peaks later at CT 20 and CT 0, respectively. CRY1 extends the repressive state into the day (CT 0-4), keeping CLOCK:BMAL in a “poised state” in absence of PER[37,113,114]. CRY2 therefore may have an important role in regulation phase by restricting CRY1s access to its DNA targets. Tissue specific effects where CRY2 plays no significant role in the peripheral liver clock are discussed[113]. Further differences between CRY1 and CRY2 are listed in Table 1. The motifs identified by Deppisch et al[115] are located at helix α10 (Figure 35) and can be used to differentiate paralogs from sequence.

Table 1: Differences of CRY paralogs. The 7M mutant substitutes seven main amino acid found in CRY2 by CRY1 amino acids.

	CRY1	CRY2	both	CRY2 7M
N-terminus		20 aa		
C-terminus	112 aa	80 aa		
Specific motif	VWPGG[115]	RLDKHL[115]		
knockout	Decreased amplitude, 1 h shorter[116,117]	Decreased amplitude, 1 h longer[116,117]	Arrhythmic [116,117]	
K_D(CLOCK) / nM	65 (6)[96]	1200 (200) [96]		130 (12) [96]
K_D(PER2 + CLOCK) / nM	196 (34) [96]	604 (29) [96]	n.a.	159 (66) [96]
Repressive phase	Early+late	early		

1.2.2.2 PER

1.2.2.2.1 Homo and heterodimerization

PER proteins can homo- and heterodimerize over their tandem PAS-A, PAS-B domains. Interestingly, different interactions sites are used in *Drosophila,* proposed as a model for insect PER homodimerization, and mammalian PERs. Importantly, heterodimerization modulate nuclear export signal (NES) function and protein stability, hence localization and half-life which are important in rhythmicity[93–95,118–120].

1.2.2.2.2 Repression by CCID

A region in the *Drosophila* PER C-terminus, the CLOCK:CYCLE interaction domain (CCID) was shown to be necessary and sufficient for repression. This region is large (270 aa, 764-1034) [121]. The CCID contains a potent nuclear localization sequence and the CK1/DBT binding domain necessary for nuclear localization as well as phosphorylation and repression of CLOCK[122,123]. The conservation of this region beyond the *Drosophila* family was not explored so far. The ability of monarch PER to repress in absence of CRY-m hints towards an at least similar region. Unfortunately, the effect of a (partial) CCID deletion was not studied in the previously mentioned monarch TRITHORAX (MLL1), CLOCK exon 19 study[78]. It remains unclear how the interaction with CLOCK is facilitated by this domain.

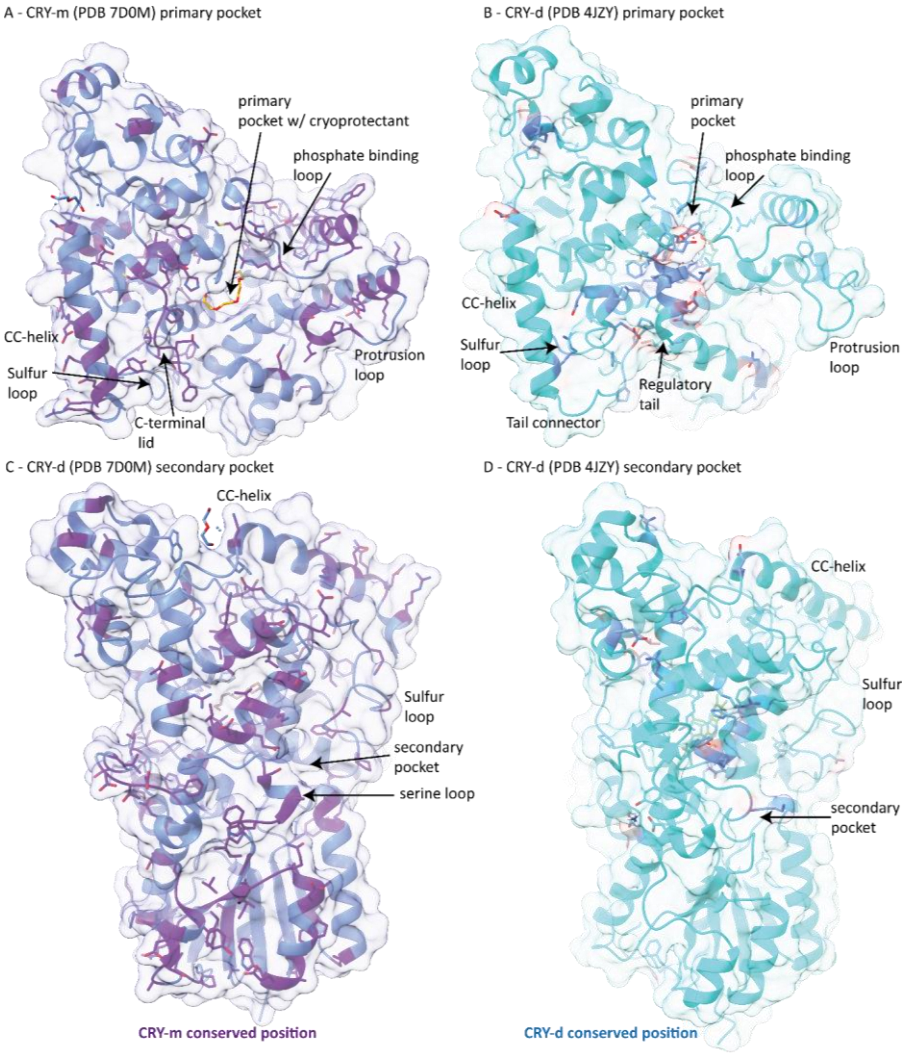

Figure 8: Differences in CRY-m and CRY-d structures and conservation. A and C show mouse CRY1 (blue) with bound cryoprotectant (orange). Positions conserved in CRY-m but not CRY-d are highlighted in purple and shown as sticks. Many areas are CRY-m specific conserved like the phosphate and protrusion loop, the C-terminal lid between CC-helix and primary pocket, residues around the secondary pocket as well as the α/β domain at the bottom in C. This structure was chosen because all loops are resolved but without a bound inhibitor.
B and D show Drosophila CRY in cyan with bound FAD (orange). CRY-d conserved positions are shown in blue and as sticks. There are much less CRY-d specific conserved positions besides the regulatory tail. FAD and regulatory tail occupy the primary pocket in CRY-d.
The solvent excluded surface is shown at high transparency. Dots indicate ordered water molecules. Based on MSA in Figure 35, Figure 36.

1.2.2.2.3 The role of TIM in repression

Drosophila TIMLESS (TIM-d) is an α-helical protein with many disordered parts between domains according to prediction and cryo electron microscopy. The α-helical domains form a continuous domain. The heterodimer with CRY-d was solved recently by cryo electron microscopy [46]. Together with PER, TIM-d forms a heterodimer, breaking the PER homodimer and stabilizing PER in the process as shown in Figure 9. TIM binds to the PAS-A, PAS-B, and cytoplasmic localization domains ($PER_{232-599}$). A triple mutation E474R/H492S/R494D on the PAS-B β-sheet surface of $PER_{232-599}$ is sufficient to break the heterodimer, but does not negatively influence the homodimer *in vitro*. Heterodimerization results in nuclear localization of PER:TIM-d. The dimerization interface used by TIM-d:PER corresponds to the mouse PER2 dimer interface[94,124,125]. No structure of the PER:TIM-d heterodimer is solved so far.

A mammalian timeless (TIM-m) homolog exists which is important in DNA replication fork protection, thereby genome stability and DNA repair. TIM-m binds to single stranded DNA via an interaction with replication protein A (RPA). TIM-m forms a complex with TIM interacting protein (TIPIN) and CRY-m[126–128], thereby connecting circadian clock and cell cycle. TIM-m is essential for development[129]. TIM-d and TIM-m arose from a gene duplication event >700 million years ago[130]

1.2.2.2.4 CK1 is necessary for repression

Probably the most important interaction of PER besides the interaction with CRY-m is the recruitment of casein kinase 1 (CK1). PER shuttles CK1 into the nucleus as they form a stable complex *in vivo* contrary to results from PER peptides[60,61,131,132]. CK1-mediated phosphorylation is an important mechanism of repression in mammals, *Drosophila* and fungi[114,122,133–136]. Loss of the CK1:PER2 interaction leads to 3 h longer periods. Expression stays rhythmic but dampened[39]. The CK1 binding sites on PER are known and conserved from insects to mammals and a functional homlog of PER in fungi[39,123,133,137–141]. CK1 phosphorylation of PER and PER associated proteins and its effects are described below.

1.2.2.2.4.1 PER phosphorylation

PER stabilization is associated with increased period while destabilization of PER2 shortens period and phosphorylation is an important determinant of PER2 stability[142]. CK1 phosphorylation sites have different effects. Phosphorylation of PER2 S478 (phosphodegron) promotes recruitment of the E3-ubiquitin ligase ß-TrCP[40,138,143,144]. Phosphorylation in the "FASP" region (starting from S662) delays degradation and decreases activity on the phosphodegron[40,131]. The phosphoswitch model proposes that a balance of phosphorylation between these two sites determines PER2 half-life[40,131].PER hyperphosphorylation increases from CT12-22 when CRY:PER:CK1 complex concentration increases at circadian promoters/enhancers, which is degraded over the ubiquitin-proteasome.

Drosophila PER stability is controlled by a similar phosphorylation system in the *per^S* (per short) and perSD (per short downstream) regions by the CK1 homolog double time (DBT). *Per^S* phosphorylation suppresses phosphorylation elsewhere in the region yielding a stable but lowly repressing PER. If *per^S* phosphorylation is blocked, perSD is phosphorylated resulting in highly active yet short-lived PER repression[140]. The phosphoswitch is an mechanism in temperature compensation[40].

1.2.2.2.4.2 Phosphorylation of PER associated proteins

CK1 phosphorylates PER1/2 associated substrates but importantly binds less to PER1[39]. CK1 phosphorylates PER-associated CLOCK leading to mobility shifts[114,145,146] and detachment from DNA[114,122]. Further, phosphomimetic mutations reduce CLK:BMAL activity[145,146] supporting the importance of phosphorylation for CLOCK repression.

1.2.2.2.4.3 DNA on/off mechanism of repression

Based on Chromatin Immuno Precipitation (ChIP) experiments in *Drosophila* it is known that PER presence at promoters results in repression ("on DNA" mechanism). CLOCK hyperphosphorylation by DBT leads to activator detachment from DNA ("off DNA") in a 1:1 PER:CLK complex in *Drosophila*[122]. This is conserved also in mammals[37,114] and *Neurospora* fungi[133,135] as shown in Figure 10.

1.2.2.2.5 PER paralogs

Three PER paralogs are present in mammals while *Drosophila* and the monarch butterfly have only one PER protein. All three paralogs in mammals contain PAS-A, PAS-B and CRY binding domains. PER1 and PER2 are important in circadian time keeping, while PER3 is mostly important in output functions and in peripheral clocks. PER1 and PER2 are partially redundant but knockout of either protein results in arrhythmicity in (PER2) or short period (PER1) in DD (Table 2).

Table 2: Differences of PER paralogs.

	PER1	**PER2**
Knockout	arrhythmic in DD[12]	arrhythmic in DD[12]
CK1 recruitment	reduced[39]	
Phosphodegron	121-126[144]	S478[139]

1.2.3 Synchronization

Synchronization or entrainment to the geophysical light-dark cycle is achieved by different means dependent on the presence of a *Drosophila*-like CRY-d. CRY-d is a photoreceptor and upon blue light exposure binds TIMELESS (TIM-d)[46,147–149] (Figure 9 A-C). The TIM-d:PER heterodimer stabilizes PER. Association of CRY-d:TIM-d results in degradation of TIM-d. The unavailability of TIM-d and thereby the lack of TIM-d stabilization shifts the balance of PER stability towards degradation[150–153] as shown in Figure 9 E. In the monarch butterfly, PER binds and stabilizes CRY-m as in mammals. Light induced TIM-d:CRY-d association and degradation leads to PER destabilization and subsequently to CRY-m destabilization[30].

Animals without a CRY-d photoreceptor use special photoreceptor cells (intrinsically photosensitive retinal ganglion cells (ipRGCs) and the light signal is transmitted neuronally to the suprachiasmatic nucleus (SCN), the central pacemaker[17–23]. The SCN uses direct efferent projections or indirectly signals via hormone production and secretion to synchronize peripheral clocks[26].

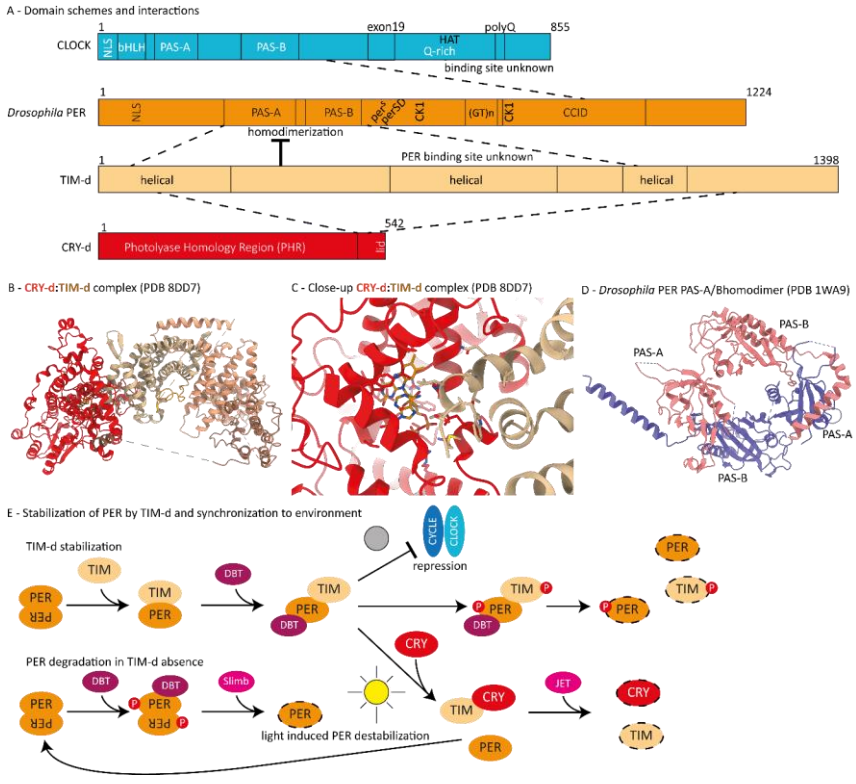

Figure 9: Repressive interactions in the Drosophila clock. A: Domain schemes and interactions (dashed lines). The PER binding sites on CLOCK and TIM-d are unknown. DBT binding sites are indicated as CK1

for simplicity. B: The structure of the CRY-d:TIM-d complex was presented in 2023 by Lin, Feng et al[46] *showing that TIM-d is helical but also largely unstructured. TIM-d is colored in multiple shades to see the different helical repeats forming a continuous domain. C: close up view of the CRY-d FAD binding pocket interaction with TIM-d N-terminus. D: Homodimerization of Drosophila PER. TIM-d binds to the PAS-A, PAS-B domains, breaking the homodimer and hindering destabilizing DBT phosphorylation. E: PER stabilization mechanism and light induced degradation by targeting the stabilizer TIM-d.*

1.2.4 TTFL models /network structures

To sum up the previous section, Figure 4 is updated to yield Figure 10 showing the interaction sites between circadian clock proteins.

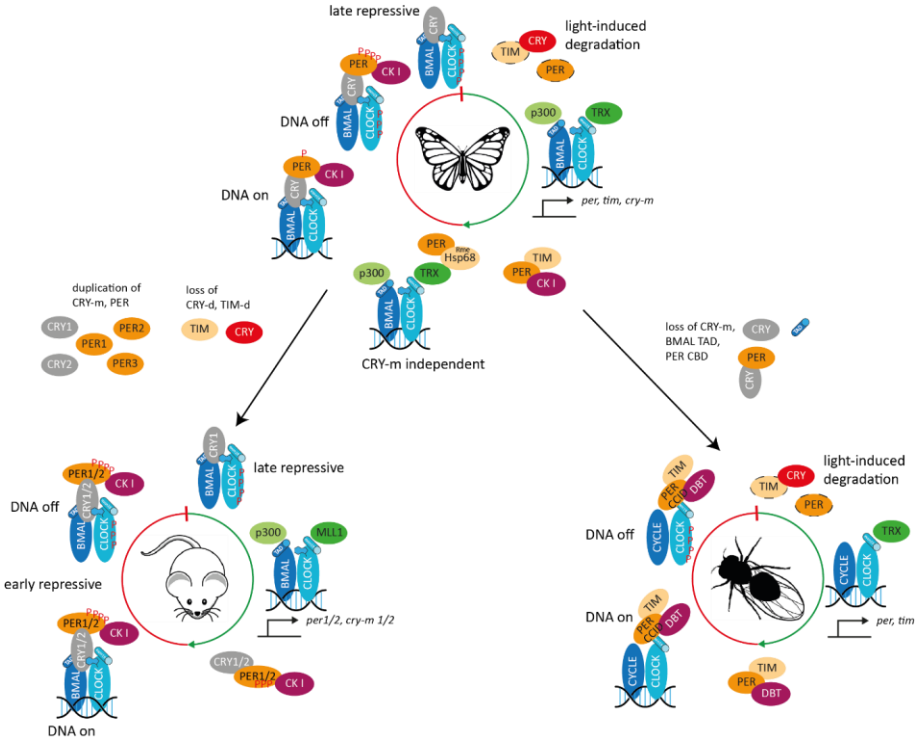

Figure 10: Mammalian and Drosophila clocks are derivatives of the ancestral clock which is conserved in the monarch butterfly. Fig 4 showed simplified clock networks, here the presented interactions and mechanisms are integrated. Evolutionary conserved functions of protein kinases within the network and repressive mechanisms are clear. The models of mouse and Drosophila are based on available literature. The model of the monarch clock is a synthesis of available literature and mouse and Drosophila models. DNA on, off repression and late repressive phase were not formally shown.

1.3 Photolyase to Cryptochrome: from UV repair to UV avoidance

CRY proteins play a central role in circadian timekeeping. Either they shuttle together with PER into the nucleus to repress circadian gene expression (CRY-m) or they are indispensable for photic entrainment (CRY-d). Interestingly, systems without photoreceptor CRY (CRY-d) or without transcription repressor CRY (CRY-m) are described. Therefore, their single functions can be dispensable if compensated for. In *Drosophila,* TIM-d is considered a replacement for CRY-m as a stabilizer and nuclear localizer. In mammals, photic entrainment is organized by neuronal pathways described above. To my knowledge no species with a functional circadian clock and without any CRY is described in detail. CRYs are therefore essential components for circadian rhythms with various, compensable functions.

The two CRYs, CRY-d and CRY-m have a common ancestor. Cryptochromes belong to the family of cryptochrome and photolyases (CRY/PL family). This family originates from ancient DNA repair enzymes the photolyases. Photolyases repair UV-induced DNA damage. Several classes of photolyases exist and are classified based on sequence identity and the types of DNA damage that they repair. Structural features and a phylogenetic tree are shown in Figure 10. The CRY-d regulatory tail in orange occupies the DNA binding site on the PHR (cyan) and after blue light illumination the TIM-d N-terminus (beige) binds there. Only in CRY-m in grey is the primary pocket not occupied by FAD but instead repurposed by FBXL3 (light coral) to control CRY-m stability. The binding site is also used by inhibitors (KL101, cyan) as described above. PER (orange) binds around the α22 helix (blue), which appears less relevant in related CRYs/PLs.

Cyclobutane pyrimidine dimer (CPD) photolyases repair cyclobutane pyrimidine dimers formed by adjacent thymine bases in the DNA which are induced by UV light. CPD PLs require blue light for cyclic electron transfer. The energy from blue light is harvested by a FAD molecule that is bound in the FAD binding pocket also called the primary pocket. CPDs can be divided into multiple classes.

(6-4) photolayses repair pyrimidine-pyrimidine (6-4) photoproducts between adjacent thymine bases. This type of DNA damage is UV-induced as well and resolved in a similar way. (6-4) PL bind FAD that works to harvest energy from blue light allowing electron transfer. An example is presented in Figure 10 in magenta showing the DNA lesion and orientation of the same towards FAD.

PLs not only bind FAD as their primary chromophore for DNA repair but also a secondary chromophore located in the secondary pocket or antenna chromophore binding pocket. These pterin or flavin derivatives transfer the excitation energy via resonance transfer to FAD[100].

CPD PLs diverged into multiple protein classes. Plant CRYs, pCRYs, work as UV-A/blue light receptors that dimerize in the dark and monomerize upon light exposure. Monomeric pCRY inactivates the E3-ubiquitin ligases CONSTITUTIVE PHOTOMORPHIC1 (COP1) and SUPPRESSOR OF PHYA-105 (SPA),

allowing their targets HY5 (ELONGATED HYPOCOTYL5) and CONSTANS (CO) to accumulate in the nucleus and activate transcription.

The last class of CRYs that I want to introduce, the CRY4-type CRYs, are also FAD containing as shown in Figure 10 in green. CRY4 are frequently found in migratory animals and seem to be involved in magnetoreception. While the theoretical quantum chemical framework is established the signaling pathways and structural organization necessary to measure magnetic inclination are still unresolved. Potential magneto-receptive functions were also discussed for *Drosophila* and monarch CRY-d[154–156].

CPD-PL and pCRYs are the most different classes, while CRY-m, (6-4) PLs and CRY-d origin from the same branch of the CRY/PL family. CRY-d branched off followed by the (6-4) PLs. The CRY4 are more closely related to (6-4) PLs than to CRY-m. Duplication of the CRY-m gene is the most recent event and gave rise to the paralogs present in some of today's animal species. Importantly the whole set of functional different genes was present in the common ancestor of animals and sets of genes were lost later in the different clades that gave rise to modern animals.

PLs are intrinsically nuclear proteins, a property that was inherited to CRY-m. The capacity to bind FAD was inherited to CRY-d. Further, the binding pockets for FAD and antenna chromophores were retained in modern CRY-m. Examples of repurposing were already presented in the previous section on interaction sites. The FAD binding/primary pocket is used by the E3-ubiquitin ligases FBXL3/21 for binding of a C-terminal tryptophan. The pocket conservation is high enough, that FAD can be forced into the pocket resulting in loss of FBXL3 binding[42]. The antenna chromophore binding pocket was repurposed for CLOCK PAS-B W362 binding[92]. Even tough CRY-m doesn't natively bind chromophores (at equimolar concentration) the chromophore binding pockets serve a function.

The interaction of TIM-d with CRY-d is also reminiscent of DNA binding. The TIM-d binding site is close to the FAD where the DNA lesion would be located in PLs. A functional commonality between CRYs is the tight regulation by or of E3-ubiquitin ligases in pCRYs and CRY-m and CRY-d.

Deppisch (2022) convincingly proposed the evolution of the CRY/PL family to origin from UV-damage mitigating enzymes to UV anticipatory proteins.

Early life evolved in the oceans. The atmosphere was lacking an ozone layer resulting in high UV radiation. Since water absorbs UV light, early life was protected from UV light at depth but any organism dwelling close to the surface for example for photosynthesis needed mechanisms to repair UV-induced DNA damage, hence CPD PLs. The ability to sense UV light and avoid it at least at its maximum intensity would have been an adaptive advantage. The blue light sensors CRY-d and pCRYs would have enabled such behaviors. A further improvement after mitigation or repair is preparedness or anticipation of predictable change, hence a circadian clock that allows to schedule behavior. Such an adaptive

mechanism is enabled by CRY-m and constant synchronization enabled by CRY-d[115]. The CRY/PL tree and representative structures are shown in Figure 11, structures are from [42,46,98,101,109,157,158].

Without a time machine we will never be able to proof this hypothesis but it is convincing considering what can be inferred from today's genes, their function, relatedness and evolutionary history.

1.4 Open questions in the circadian central loop

Concluding from the vast body of knowledge presented above, some key interactions remain elusive.

- How do *Drosophila* CLOCK:CYCLE achieve transactivation in the absence of a BMAL TAD?
- How does the PER CCID bind to CLOCK:CYCLE and is DBT activity the only mode of repression?
- What is the structure of *Drosophila* PER:TIM-d complex?
- What are the structures of the CLOCK:CRY complex or of the BMAL1:CLOCK:CRY1 poised state?
- How does CLOCK exon19 interact with MLL1?
- Can mammalian PER repress CLOCK without CRY-m as in monarch and *Drosophila*?
- Does the BMAL TAD really bind the CRY CC-helix and if so, how can PER binding have no influence?
- Are there species with circadian clocks lacking CRY-d and CRY-m?
- What is the adaptive advantage of maintaining multiple PER and CRY paralogs?

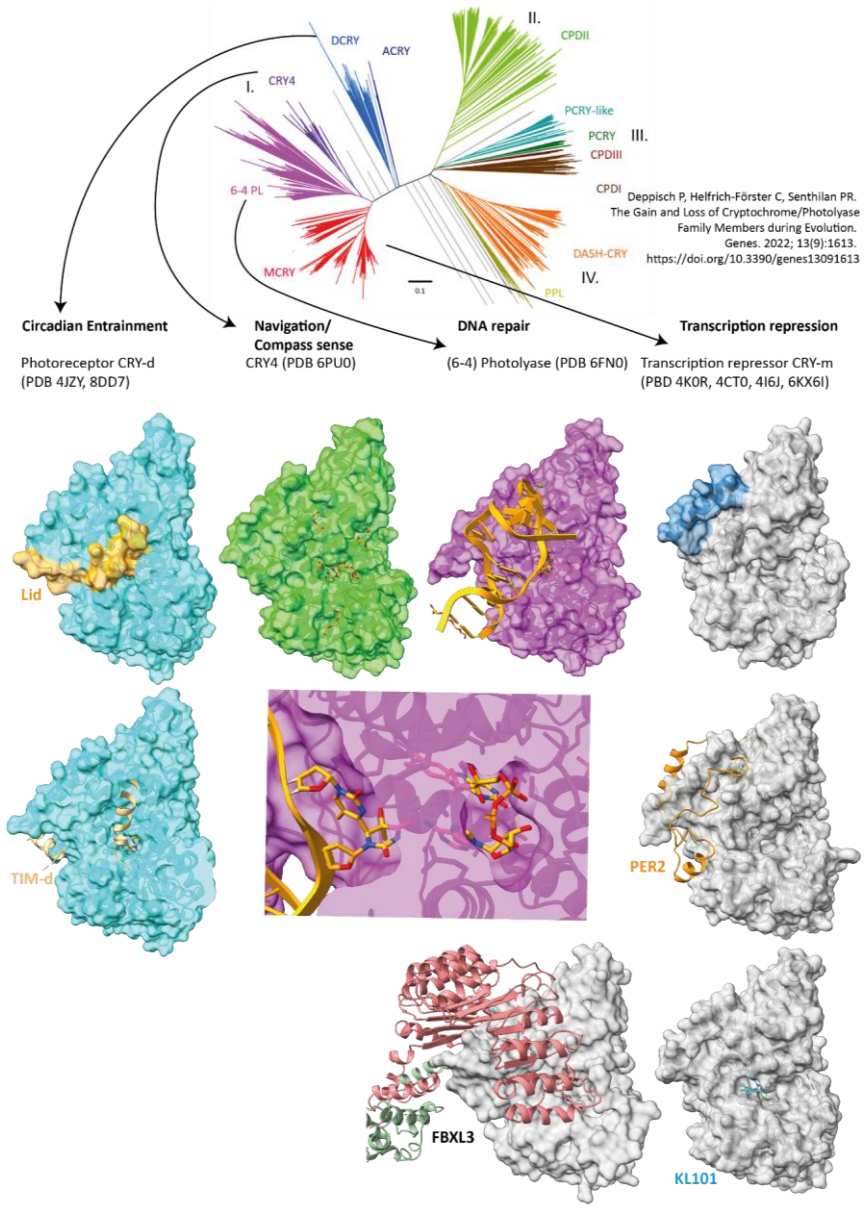

Figure 11: Evolution of the CRY/PL family. The unrooted phylogenetic tree of the CRY/PL family is shown from Deppisch et al (2022)[115], Fig. 2. Representative crystal structures of the family members demonstrate the functional and structural diversity in the CRY/PL family. All CRY/PLs besides the transcription repressors bind FAD (carbons in orange) for either energy harvesting (DNA repair, magnetic sense) or photoreception. Importantly, the FAD binding pocket is repurposed for circadian time-keeping as shown by the FBXL3 interaction and the effect of inhibitors.

1.5 Aim of the thesis

1.5.1 Identify the BMAL TAD binding site on CRY-m

The current model of interaction of the CRY-m CC-helix with the BMAL TAD is inconclusive. The underlying data can be explained with an association of the isolated CRY-m coiled-coil helix forming a coiled-coil with the BMAL1 TAD α-helix[71,83]. These experiments were informed by results showing the necessity of the CC-helix for repression[97]. A fact that was later explained by the binding of PER to the CC-helix[98,99] which is necessary for strong repression. The CC-helix is an integral part of the CRY-m surface and cannot easily be compared to a monomeric peptide of it in solution. The binding site could be anywhere on the CRY-m PHR and identification of this important interaction site awaits.

1.5.2 Investigate the PER/BMAL competition for CRY-m binding under the current model

A strong competition is suggested of BMAL TAD and PER CBD for CRY-m binding under the current model of BMAL:CRY-m interaction[71] and the later presented CRY-m:PER crystal structures[98,99]. The PER CBD wraps tightly around the CC-helix further limiting the available solvent-accessible surface. But data was presented that showed unperturbed binding of the BMAL TAD to apo CRY1/2 PHR and CRY1/2:PER2[96]. This can be explained by a different BMAL TAD binding site. These interactions require further investigation.

1.5.3 Investigate evolution of the clock network and its defining interaction sites

Three general network structures for the circadian clock were proposed. Identification of those networks was based on experimental testing of the cryptochromes present in the species for repression and photoreception[43]. I hypothesize that neither experimentally determined CRY function nor sequence is required to unambiguously identify the network structure. Instead BMAL or PER sequences together with TIM-d presence can be used. I further hypothesize that all interactions between CRY-m and its interactors are inherited or lost together.

Furthermore, the distribution of these clock networks in the insect class should be determined. Current evolutionary studies focus on gene presence, but I propose that interaction site presence would be the better proxy for function and conservation of the interaction sites and should therefore be investigated.

Lastly, I aim to identify the adaptive advantage of maintaining multiple CRY-m and PER genes and how flies are doing the trick of not using a CRY-m.

2 Materials

Table 3:Buffers

Name	Composition	Stock	from
5xSDS	62.5 mM Tris-HCl pH 6.8, 2% SDS, 25% glycerol, 0.01% bromophenol blue, 5% BME		
1x TAE	40 mM Tris, 20 mM Acetic acid, 1 mM EDTA	50x TAE	IMB media lab
1x MOPS SDS-PAGE running buffer	50 mM MOPS, 50 mM Tris, 5 mM EDTA, 0.5 % SDS		
Coomassie Simple stain	6.7% (v/V) ethanol, 4.4% (v/v) acetic acid, 10 g/L β-cyclodextrin, 115 mg/L Coomassie blue G-250		
Tbf1	30 mM Kac, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% (v/v(glycerol, pH 5.8 by Hac, sterile filtered		
Tbf2	10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$, 15% (v/v) glycerol, pH 5.8 by Hac/KOH, sterile filtered		
Transfer buffer	144.1 g/L glycine, 30.3 g/L Tris	10x	IMB media lab
TBS Buffer	10 mM Tris pH 7.5, 150 mM NaCl		
TBS-Tween/Triton Buffer	20 mM Tris pH 7.5, 500 mM NaCl, 0.05% Tween 20, 0.2% Triton-X-100		

Table 4: Media

Name	Composition/Order nr.	from
TB	Tryptone 12 g/L, yeast extract 24 g/L, glycerol 4 mL/L	Prepared by IMB media lab
10x TB salts	0.17 M KH_2PO_4, 0.72 M K_2HPO4	Prepared by IMB media lab
IceT	80% black tea, 20% juice	
LB	10 g/L NaCl, 10 g/L trypton, 5 g/L yeast extract, pH 7.0	Prepared by IMB media lab
LB-agar	15 g/L agar-agar to LB	Prepared by IMB media lab
LB-agar Amp	Add 100 µg/mL ampicillin to LB-agar	Prepared by IMB media lab
LB-agar Kan	Add 30 µg/mL kanamycin to LB-agar	Prepared by IMB media lab
LB-agar Kan/Tet/Gent/IPTG/X-Gal	Add 50 µg/mL Kanamycin, 10 µg/mL Gentamycin, 10 µg/mL Tetracyclin,	

Name	Composition/Order nr.	from
	100 µg/mL X-Gal in DMSO, 40 µg/mL IPTG	
SF900 III	12658027	Gibco™
SF900 II Combo Kit	Discontinued, 10902096	Prepared by IMB media lab

Table 5:Purification buffers

Buffer	Composition
HisA Tris	50mM Tris, 300 mM NaCl, 10% (v/v) glycerol, 5 mM β-Mercaptoethanol, pH 7.5
HisB-Tris	50mM Tris, 300 mM NaCl, 300 mM imidazole, 10% (v/v) glycerol, 5 mM β-Mercaptoethanol, pH 7.5
HisA-SPG	50mM SPG, 300 mM NaCl, 10% (v/v) glycerol, 5 mM β-Mercaptoethanol, pH 7.5
HisB-SPG	50mM SPG, 300 mM NaCl, 300 mM imidazole, 10% (v/v) glycerol, 5 mM β-Mercaptoethanol, pH 7.5
GSH-lysis buffer	50mM Tris, 200 mM NaCl, 5% (v/v) glycerol, 1 mM EDTA, 14 mM β-Mercaptoethanol, pH 7.5
GSH-Wash	50mM Tris, 1 M NaCl, 5% (v/v) glycerol, 50 mM KCl, 10 mM $MgCl_2$, 10 mM ATP 14 mM β-Mercaptoethanol, 1 µL/10 mL Sm nuclease, pH 7.5
GSH-E	50mM Tris, 200 mM NaCl, 5% (v/v) glycerol, 1 mM EDTA, 14 mM β-Mercaptoethanol, 20 mM GSH reduced, pH 7.5
GST-SPG-Wash	50mM SPG, 1 M NaCl, 5% (v/v) glycerol, 50 mM KCl, 10 mM $MgCl_2$, 10 mM ATP 14 mM β-Mercaptoethanol, 1 µL/10 mL Sm nuclease, pH 7.5
GST-SPG-E	50mM SPG, 200 mM NaCl, 5% (v/v) glycerol, 1 mM EDTA, 14 mM β-Mercaptoethanol, 20 mM GSH reduced, pH 7.5
IEX-A-Tris	50mM Tris, 10% (v/v) glycerol, 10 mM β-Mercaptoethanol, pH 7.5
IEX-B-Tris	50mM Tris, 1 M NaCl, 10% (v/v) glycerol, 10 mM β-Mercaptoethanol, pH 7.5
IEX-A-SPG	50mM SPG, 10% (v/v) glycerol, 10 mM β-Mercaptoethanol, pH 6.5
IEX-B-SPG	50mM SPG, 1 M NaCl, 10% (v/v) glycerol, 10 mM β-Mercaptoethanol, pH 6.5
SEC buffer	25 mM HEPES, 150 mM NaCl, 5% (v/v) glycerol, 0.5 mM TCEP, pH 7.5
SEC buffer SPG	25 mM SPG, 150 mM NaCl, 5% (v/v) glycerol, 0.5 mM TCEP, pH 7.5

Table 6: Equipment

Equipment	Specification	Manufacturer
Agarose gel electrophoresis power supply	Power Pack P25 T	Biometra ®
Blue light transilluminator	TruBlu 2	EDVOTEK®
Centrifuges	Avanti JXN-26 Heraeus Fresco 21 Eppendorf 5415 C Eppendorf 5810 R	Beckman Coulter Hettich Lab Technology Eppendorf AG ThermoFisher Scientific Inc.
Cooling cabinet	Unichromat	UniEquip
FPLC systems	NGC™ chromatography system Äkta Prime Plus	Biorad Äkta, Cytiva
Freezer (-80 °C)	Panasonic VIP Plus	Panasonic
Freezer (-20 °C)	Liebherr MEDline	Liebherr International AG
Fridge (4 °C)	Liebherr MEDline	Liebherr International AG

Equipment	Specification	Manufacturer
SDS-PAGE Gel chamber	BioRad Mini-PROTREAN Tetra System	BioRad Laboratories GmbH
Western Blot transfer cell	Trans-Blot® Semi-Dry	Bio-Rad
Agarose gel imaging system	Odyssey ® FC	LI-COR Biosciences
SDS gel scanner	Scanjet 300	HP Deutschland GmbH
Heat Block	MY Block™ Mini Dry Bath Accublock™ Mini Thermomixer 5436	Benchmark Labnet international Inc. eppendorf
Incubation shaker	INFORS CH-4103 Bottmingen (bacteria cells) INFORS HAT Multitron Pro (insect cells)	InforsAG
Laminar flow hood	BSC-SG403EN	Dometic
Microscale	PRACTUM224-1S®	Sartorius
Precision scale	L420D	Sartorius
Microwave	NN-GT462M	Panasonic
Spectrophotometer	Nanodrop DS-11 + Spectrophotometer	DeNovix
PCR cycler	TProfessional Primus	Biometra® MWG-Biotech
pH-meter	Lab 850	SI Analytics
Plate reader	Spark 20M	Tecan Trading AG
Peristaltic pump	Peristaltic Pump P-1	Cytiva
Rotors	75003424 A-4-81 F-45-18-11 JA-25.50 JLA-8.100	Hettich Lab Technology ThermoFisher Scientific Inc. eppendorf Beckmann Coulter
SDS-PAGE power supply	Power Pack 300	BioRad Laboratories GmbH
Sonifier	Analog Sonifier Modell 450-CE 7/2 Tapped 12,7mm	Branson Ultrasonics
Vortexer	REAX 2000 Vortex Genie 2	Heidolph Instruments GmbH & Co. KG Scientific Industries Inc.
UV Crosslinker 365 nm	1777.1	Carl Roth, @ IMB König Lab
ITC	MicroCal PEAQ-ITC	Malvern Panalytical Ltd

Table 7: Special Chemicals and Consumables

Product	Name/Order #	Manufacturer
Cellfectin II	10362100	Gibco™
Protease inhibitor cocktail	cOmplete™, EDTA-free Protease inhibitor cocktail	Roche Diagnostics GmbH
Sulfo-SDA	26173	Thermo Scientific™
BS3	21580	Thermo Scientific™
FP plates	3544	Corning
96-well PCR plates	Mikrotestplatte 96 Well, F	Sarstedt AG & Co. KG
Centrifugal filters	Amicon®Ultra-15 Centrifugal Filters	Merck Milipore
Dialysis tube	MembraCel cellulose (MWCO 14,000)	Carl Roth GmbH+ Co.KG

Table 8: Western blot

Name	Order nr.	Manufacturer
His_5-HRP Conjugate Kit	34460	Qiagen
HRP substrate	RPN2106	Amersham
Membrane	Immobilon-P	Merck
Whatman Paper	CL75.1	Roth

Table 9: RNA work

Product	Order number/composition	Manufacturer
Oligo (dT) 12-18	18418012	Invitrogen
Random hexamer	N8080127	Invitrogen
RNaseOut	10777019	Invitrogen
Glycogen, RNA grade	R0551	Thermo Scientific™
Trizol	15596026	Invitrogen
5x first-strand buffer	250 mM Tris, 375 mM KCl, 15 mM $MgCl_2$, pH 8.0	IMB core facility
RT enzyme	400 U/µL	IMB core facility

Table 10: Columns and affinity resins

Column/resin	Product name	Manufacturer
Affinity chromatography		
HisTrap	HisTrap FF crude 5mL	Cytiva
Ni-NTA beads	Ni-NTA His-Bind® Superflow™ Resin	EMD Millipore Corp.
Magnetic Ni-NTA beads	Dynabeads™ His-Tag Isolation & Pulldown	Invitrogen
GST mag	Pierce™ Glutathione Magnetic Agarose Beads	Thermo Scientific
GST agarose	Pierce™ GST agarose	Thermo Scientific
Ion-exchange chromatography	HiTrap SP HP 5mL	Cytiva
	HiTrap Q HP 5mL	Cytiva
Size-exclusion	S200 10/300 increase	Cytiva
	S200 16/600	Cytiva
	S75 10/300 increase	Cytiva
	S75 16/600	Cytiva
DNA		
Mini prep columns	NucleoSpin® Plasmid columns	MACHERY-NAGEL GmbH & Co. KG
PCR cleanup columns	NucleoSpin® Gel and PCR clean-up columns	MACHERY-NAGEL GmbH & Co. KG

Table 11: Commercial Kits

Name	Order nr.	Manufacturer
RNeasy Mini Kit	74104	Qiagen
QIAprep Spin Miniprep Kit	27104	Qiagen

Table 12: Standards

Name	Order nr	Manufacturer
1 kb DNA Ladder	Y014.2	Carl-ROTH GmbH
Protein Ladder		IMB PPCF
Roti®-Mark TRICOLOR	8271.2	Carl-ROTH GmbH
SYBR Safe	S33102	Invitrogen

Chemicals and consumables were purchased from Applichem (Darmstadt, Germany), Carl Roth GmbH & Co. KG (Karlsruhe, Germany), Corning GmbH (Karlsruhe, Germany), GE Healthcare (München, Germany), Qiagen (Hilden, Germany), Macherey-Nagel (Düren, Germany), Merck KGaA (Darmstadt, Germany), Roche Diagnostics (Mannheim, Germany), Sartorius Stedim Biotech (Göttingen, Germany), Sarstedt AG (Nümbrecht, Germany), SERVA Electrophoresis (Heidelberg, Germany), Sigma Aldrich (München, Germany) and VWR International.

Table 13: Peptides

No	Name	Sequence (N – C-terminus)	MW
1	FAM-mBMAL-TAD-Cterm	gDFSDLPWPL-	1504.0
2	FAM-inBMAL-TAD-Cterm	gDFSGLPWPLP	1543.0
3	FAM-lepBMAL-TAD-Cterm	gNFSGLPWPLP	1542.0
4	FAM-dpBMAL-TAD-Cterm	gNISGLPWPLP	1508.0
5	FAM-lep-in_corrBMAL-TAD-Cterm	gNFSDLPWPL-	1503.0
6	FAM-dp-in_corrBMAL-TAD-Cterm	gNISDLPWPL-	1469.0
7	FAM-in_GD-BMAL-TAD-Cterm	gDFSDLPWPLP	1601.0
8	FAM-in_P--BMAL-TAD-Cterm	gDFSGLPWPL-	1446.0
9	FAM-m_FI-BMAL-TAD-Cterm	gDISDLPWPL-	1470.0
10	FAM-mBMAL-TAD-alpha	SNDEAAMAVIMSLLEAD	2137.0
11	FAM-iBMAL-TAD-alpha	GNDEAAMAVIMSLLEAD	2107.0
12	FAM-dpBMAL TAD	GNGEAAMAVIMSLLEADAGLGGPVNISGLPWPLP	3677.1
13	FAM-mBMAL-TAD	SNDEAAMAVIMSLLEADAGLGGPVDFSDLPWPL-	3761.1
14	FAM-dpBMAL TAD-alpha	GNGEAAMAVIMSLLEAD	1876.9
15	FAM-dpBMAL-TAD-woalpha	GN----------AGLGGPVNISGLPWPLP	2173.3

All peptides are also available without the N-terminal FAM label (free amine). The lower case g denotes a spacer glycine included in the C-term peptides. All peptides besides the TAD peptides were FAM labeled by Dr. Fabian Barthels. mBMAL-TAD (#13), mBMAL-TAD-alpha (#10), mBMAL-TAD-Cterm (#1) and dpBMAL-TAD-Cterm (#4) were also modified with a N-terminal trifluoromethyl phenyl diazirine (TPD) moiety for photoaffinity labeling by Dr. Fabian Barthels.

Table 14:Plasmid backbones

Vector	Fusion Tag	Selection Antibiotic	Cleavage site	Host
pCoofy1	N-term 6xHis	Kanamycin	3C	*E.coli*
pCoofy3	N-term 6xHis-GST	Kanamycin	3C	*E.coli*
pCoofy4	N-term 6xHis-MBP	Kanamycin	3C	*E.coli*
pCoofy6	N-term 6xHis-SUMO	Kanamycin	Sumo	*E.coli*
pET50b	N-term-6xHis-NusA-6xHis	Kanamycin	3C	*E.coli*
pEC 3C His	N-term-6xHis	Kanamycin	3C	*E.coli*
pGFP	N-term-6xHis-twinStrep-3C-GFP	Ampicillin	3C	Insect cells
pBig1a	Depends on inserts	Ampicillin Streptomycin	3C	Insect cells
pCoofy27	N-term 6xHis	Ampicillin	3C	Insect cells
pCoofy28	N-term 6xHis-GST	Ampicillin	3C	Insect cells
pFastBac HT B	N-term 6xHis	Ampicillin	TEV	Insect cells
pFastBac dualmod	6xHis-C-term	Ampicillin	3C	Insect cells
pDEST26	N-term 6xHis	Ampiclin		Mammalian

Table 15: Software

Software	Manufacturer/Consortium
Jalview[159]	ELIXIR-UK resource
ChimeraX[160,161]	RBVI, UCSF
PyMOL	Schroedinger
GraphPad Prism	Dotmatics
Illustrator	Adobe
SnapGene	Dotmatics
AlphaFold	Google DeepMind
Word	Microsoft
Excel	Microsoft
Mendeley	Elsevier
MicroCal PEAQ-ITC Analysis Software	Malvern Panalytical Ltd

Table 16: Webtools

Tool	URL
Clustal[162]	https://www.ebi.ac.uk/Tools/msa/clustalo/
Timetree[163]	http://www.timetree.org/
iTOL[164]	https://itol.embl.de/
BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi
ESPript3.0[165]	https://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi
Protparam[166]	https://web.expasy.org/protparam/
Weblogo[167,168]	https://weblogo.berkeley.edu/logo.cgi

Table 17: Primers

#	Primer	Seq 5' -3'
2	dpPER-ccdB-R	CCCCAGAACATCAGGTTAATGGCGctaagaatctgatgtaccggcaacgttatctg
5	dpPER_N926-3C-F	AAGTTCTGTTCCAGGGGCCCaacaagagccaggcacaacagaaaaagg
6	dpPER_N926-SUMO3-F	GTGTTCCAGCAGCAGACCGGTGGAaacaagagccaggcacaacagaaaaagg
17	ccdB	CGCCATTAACCTGATGTTCTGGGG
18	3C	GGGCCCCTGGAACAGAACTTCCAG
19	Sumo3	TCCACCGGTCTGCTGCTGGAACAC
32	dpCRY2-stop-R	GCGGCCGCGACTAGTGTTAatttttctttttccgatgagatcttatctgtttgtgg
33	tev-dpCRY2-F	GAAAACCTGTATTTTCAGGGCGCCatgtcggttgccgagacg
34	mCRY2-3C-F	GGGCCCCTGGAACAGAACTTCCAGatggcggcggctgc
35	mCRY2-512-ccd-R	CGCCATTAACCTGATGTTCTGGGGtcagtatctcgacagctgttggtag
40	mCRY2-PHR-F	CTGGAAGTTCTGTTCCAGGGGCCCatggcggcggctgc
41	mCRY2-PHR-R	CCCCAGAACATCAGGTTAATGGCGtcagtatctcgacagctgttggtag
42	dpCRY2-3C-F	AAGTTCTGTTCCAGGGGCCCatgtcggttgccgagacg
43	dpCRY2-PHR-R	CCCCAGAACATCAGGTTAATGGCGtcatttaaacttcgctagttgtgcgtacacc
47	dpCRY2-ccdB	CCCCAGAACATCAGGTTAATGGCGtcaatttttctttttccgatgagatcttatctgtttgtg
62	KanR-F	caccgctgagcaataactatcataaccccttgg
63	KanR-R	aaaggcggtaatacggttatcc

#	Primer	Seq 5’ -3’
65	3C-mPER1-F	AAGTTCTGTTCCAGGGGCCCgaccaggtcattaagtgtgtgctcc
66	SUMO3-mPER1-F	GTGTTCCAGCAGCAGACCGGTGGAgaccaggtcattaagtgtgtgctcc
67	ccdB-mPER1-R	CCCCAGAACATCAGGTTAATGGCGTCAttgaacgctgctgccacagtc
70	3C-mPER1-S1013-F	AAGTTCTGTTCCAGGGGCCCagtgaggagactgctgagcc
71	3C-mPER1-S1038-F	AAGTTCTGTTCCAGGGGCCCtccagcgacctgctggag
72	3C-mPER1-S1079-F	AAGTTCTGTTCCAGGGGCCCagcacctcagccagcatc
73	3C-mPER1-Q1089-F	AAGTTCTGTTCCAGGGGCCCcagagcagccatacaagcaagtac
74	3C-mPER1-S1102-F	AAGTTCTGTTCCAGGGGCCCtcttccgaggctgaagctggg
75	3C-mPER1-P1116-F	AAGTTCTGTTCCAGGGGCCCcctggggaccaggtcattaagtg
76	ccdB-mPER1-P1229-R	CCCCAGAACATCAGGTTAATGGCGTCAgggctccagccccaatcc
77	ccdB-mPER1-G1241-R	CCCCAGAACATCAGGTTAATGGCGTCAaccacacccaccaccctc
78	ccdB-mPER1-K1260-R	CCCCAGAACATCAGGTTAATGGCGTCActtagccccaatttgggtctggg
79	ccdB-mPER1-S1291-R	CCCCAGAACATCAGGTTAATGGCGTCAgctggtgctgttttcttctgcag
124	124_pDEST-F	taAAGGGTGGGCGCGCCGAC
125	125_pDEST-R	AGCCTGCTTTTTTGTACAAACTTGtg
127	127_agCRY2_D961-R	CCCCAGAACATCAGGTTAATGGCGCTAATCCCCCGCACCAATC
128	128_agCRY2_M1-F	CTGGAAGTTCTGTTCCAGGGGCCCATGCGGGACAAACACAC
129	129_agCRY2_Y494-R	CCCCAGAACATCAGGTTAATGGCGTTAGTATTTGGCAAGATGTTGGTACAC
130	130_amCRY2_G20-F	CTGGAAGTTCTGTTCCAGGGGCCCGGAAAACACACGGTCCATTG
131	131_amCRY2_H570-R	CCCCAGAACATCAGGTTAATGGCGTTAATGATGATGCTGTCGTTGC
132	132_amCRY2_M1-F	CTGGAAGTTCTGTTCCAGGGGCCCATGACGGGTAGTCGTAGTAGC
133	133_amCRY2_Y511-R	CCCCAGAACATCAGGTTAATGGCGTTAATATTTATTTAATTGTTGATAAACTTGTTTCATTCTTTC
134	134_tcCRY2_D16-F	CTGGAAGTTCTGTTCCAGGGGCCCGATAAACATATGGTACACTGGTTCC
135	135_tcCRY2_K535-R	CCCCAGAACATCAGGTTAATGGCGTTATTTCCTACTGGGGTTACCAAC
136	136_tcCRY2_M1-F	CTGGAAGTTCTGTTCCAGGGGCCCATGAGTGGGGTGGTTGG
137	137_tcCRY2_Y507-R	CCCCAGAACATCAGGTTAATGGCGTTAGTAGTTGGAGAGCTGTTGATACAC
138	138_tnCRY2_G22-F	CTGGAAGTTCTGTTCCAGGGGCCCGGGAAGCACACGGTTCAC
139	139_tnCRY2_M1-F	CTGGAAGTTCTGTTCCAGGGGCCCATGACGGGAAGTACTAACAACG
140	140_tnCRY2_Q573-R	CCCCAGAACATCAGGTTAATGGCGCTATTGTTGCTGTTGTTGCTG
141	141_tnCRY2_Y513-R	CCCCAGAACATCAGGTTAATGGCGTTAATACTTATTCAGCTGCTGGTACAC
142	142_aaCRY2_D822-R	CCCCAGAACATCAGGTTAATGGCGCTAATCACCGGACGTTAGTTTTTC
143	143_aaCRY2_M1-F	CTGGAAGTTCTGTTCCAGGGGCCCATGGGAATGACCAAGCAG
144	144_aaCRY2_Q44-F	CTGGAAGTTCTGTTCCAGGGGCCCCAACAGAAGCACACCGTT
145	145_aaCRY2_Y536-R	CCCCAGAACATCAGGTTAATGGCGTTAATATTTGGCCAACTGCTGG
146	146_mPER1_dCtermSTOP-F	cagcgttcaagatcctggcTAAgatagtgactggatatgttgtgttttac
147	147_mPER1_dCtermSTOP-R	gtaaaacacaacatatccagtcactatcTTAgccaggatcttgaacgctg
150	150_mPER2-stop-F	GTGAGGAGAAAGGCAACATTTAAATAGTGACTGGATATGTTGTG
151	151_mPER2-stop-R	CACAACATATCCAGTCACTATTTAAATGTTGCCTTTCTCCTCAC
154	154_mPER2_dCterm-F	GAGAAAGGCAACATTTAAAAGGGTGGGCGCGCCG
155	155_mPER2_dCterm-R	CCCTTTTAAATGTTGCCTTTCTCCTCAC
158	158_mPER1-E1105-core-R	ctcggaagagtcgatgctg

#	Primer	Seq 5’ -3’
159	159_mPER2-T1257-tail-R	GTCGGCGCGCCCACCCTTTACGTCTGGGCCTCTATC
160	160_mPER2-S1215-CBD-R	GTCGGCGCGCCCACCCTTtaACTCTCGCAGTAAACACAG
161	161_mPER2-core-R	CTCTGAAGAGTCAATGCTTCC
162	162_mPER1-A1106-CBD-F	GGAAGCATTGACTCTTCAGAGgctgaagctggggctg
163	163_mPER2-N1113-CBD-F	cagcatcgactcttccgagAATAATCACAAAGCAAAAATGATCCC
164	164_mPER1_Q1210_CBD-R	GTCGGCGCGCCCACCCTTtattgaacgctgctgccacag
165	165_mPER1-S1292-CBD+tail-R	CGCGCCCACCCTTtag
166	166_mPER1_Q1210-R	ttgaacgctgctgccacag
167	167_mPER2-E1216-tail-F	ctgtggcagcagcgttcaaGAGGAGAAAGGCAACATTTGC
168	168_mPER1-S1208-tail-F	agcgttcaagatcctggc
169	169_mPER2-S1215-CBD-R	gccaggatcttgaacgctACTCTCGCAGTAAACACAG
170	170_CRY1_K151A-F	GACAGCCACCTCTAACATATGCAAGGTTTCAGACTCTCGTC
171	171_CRY1_K151A-R	GACGAGAGTCTGAAACCTTGCATATGTTAGAGGTGGCTGTC
172	172_CRY1_K159A-F	GTTTCAGACTCTCGTCAGCGCGATGGAGCCACTGGAG
173	173_CRY1_K159A-R	CTCCAGTGGCTCCATCGCGCTGACGAGAGTCTGAAAC
174	174_CRY1_R227A-K228A-F	CGTTTGGAAAGGCATTTGGAAGCGGCAGCCTGGGTGGCAAAC
175	175_CRY1_R227A-K228A-R	GTTTGCCACCCAGGCTGCCGCTTCCAAATGCCTTTCCAAACG
176	176_CRY1_R236A-R238A-F	CTGGGTGGCAAACTTTGAAGCCCCTGCAATGAATGCAAACTCCCTG
177	177_CRY1_R236A-R238A-R	CAGGGAGTTTGCATTCATTGCAGGGGCTTCAAAGTTTGCCACCCAG
178	178_CRY1_H355A-F	GGAGGGCTGGATCCACGCATTAGCCAGACACGCGG
179	179_CRY1_H355A-R	CCGCGTGTCTGGCTAATGCGTGGATCCAGCCCTCC
180	180_CRY1_Q407E-Q408E-F	CTGTCCTGCAGTTCCTTTTTTGAAGAATTTTTTCACTGCTACTGCCC
181	181_CRY1_Q407E-Q408E-R	GGGCAGTAGCAGTGAAAAAATTCTTCAAAAAAGGAACTGCAGGACAG
182	182_CRY1_H411A-F	CAGTTCCTTTTTTCAGCAATTTTTTGCCTGCTACTGCCCTGTGG
183	183_CRY1_H411A-R	CCACAGGGCAGTAGCAGGCAAAAAATTGCTGAAAAAAGGAACTG
184	184_CRY1_R421A-F	CCTGTGGGTTTTGGTAGGGCGACAGATCCCAATGGAGAC
185	185_CRY1_R421-R	GTCTCCATTGGGATCTGTCGCCCTACCAAAACCCACAGG
186	186_mCRY1-PHR-R	TAGCCCTCTGTACCGGG

Table 18: Enzymes

Enzyme	Purchased from
His-3C Protease	IMB Core facility
GST-3C Protease	IMB Core facility
Sm nuclease	IMB Core facility
HF-DNA Polymerase	IMB Core facility
Screen-Blend DNA Polymerase mix	IMB Core facility
2xGibson assembly mix	IMB Core facility

Table 19: CRY Proteins

CRY	ID	fragment	tags
mCRY1 fl	P97784	1-608 WT	C-term 6xHis
mCRY1 fl	P97784	1-608 K151A	N-term 6xHis pDEST26
mCRY1 fl	P97784	1-608 K159A	N-term 6xHis pDEST26
mCRY1 fl	P97784	1-608 R227A, K228A	N-term 6xHis pDEST26
mCRY1 fl	P97784	1-608 R236A, R238A	N-term 6xHis pDEST26
mCRY1 fl	P97784	1-608 H355A	N-term 6xHis pDEST26
mCRY1 fl	P97784	1-608 Q407E, Q408E	N-term 6xHis pDEST26
mCRY1 fl	P97784	1-608 H411A	N-term 6xHis pDEST26
mCRY1 fl	P97784	1-608 K421A	N-term 6xHis pDEST26
mCRY1 PHR	P97784	1-496 WT	N-term 6xHis 3C
mCRY1 PHR	P97784	1-496 K151A	N-term 6xHis 3C
mCRY1 PHR	P97784	1-496 K159A	N-term 6xHis 3C
mCRY1 PHR	P97784	1-496 R227A, K228A	N-term 6xHis 3C
mCRY1 PHR	P97784	1-496 R236A, R238A	N-term 6xHis 3C
mCRY1 PHR	P97784	1-496 H355A	N-term 6xHis 3C
mCRY1 PHR	P97784	1-496 Q407E, Q408E	N-term 6xHis 3C
mCRY1 PHR	P97784	1-496 H411A	N-term 6xHis 3C
mCRY1 PHR	P97784	1-496 K421A	N-term 6xHis 3C
mCRY2 fl	Q9R194	1-592	C-term 6xHis
mCRY2 PHR	Q9R194	1-512	N-term 6xHis 3C
aaCRY2 PHR	XM_001655728.2	44-536	N-term 7xHis 3C
agCRY2 fl	DQ219483.1	1-961	N-term 7xHis 3C
agCRY2 PHR	DQ219483.1	1-494	N-term 7xHis 3C
amCRY2 fl	EF117814.1	1-570	N-term 7xHis 3C
amCRY2 PHR	EF117814.1	1-511	N-term 7xHis 3C
tcCRY2 PHR	EF117815.1	1-507	N-term 7xHis 3C
dpCRY2 fl	Q0QWP3	1-742	N-term 7xHis 3C
dpCRY2 PHR	Q0QWP3	1-520	N-term 7xHis 3C
dpCRY2 PHR	Q0QWP3	1-520	N-term 7xHis 3C-GST
dpCRY2 PHR	Q0QWP3	1-520	N-term 6xHis-twinStrep-3C-GFP-3C

Table 20: PER proteins

PER	ID	PER1 fragment	PER2 frag	tags
PER1	O35973	1116-1291		N-term 6xHis-GST-3C
PER1 CBD+tail	O35973	1109-1241		N-term 6xHis-GST-3C
PER1 CBD	O35973	1109-1209		N-term 6xHis-GST-3C
PER2 CBD+tail	O54943		1119-1252	N-term 6xHis-3C
PER2 CBD	O54943		1119-1215	N-term 6xHis-3C
PER1 WT	O35973	1-1291		N-term 6xHis pDEST26
PER1 ΔCterm	O35973	1-1212		N-term 6xHis pDEST26
PER1 core PER2 CBD	O35973	1-1104	1113-1215	N-term 6xHis pDEST26
PER1 core PER2 CBD+tail	O35973	1-1104	1113-1257	N-term 6xHis pDEST26
PER1 core PER2 CBD PER1 tail	O35973	1-1104, 1207-1291	1113-1215	N-term 6xHis pDEST26
PER2 WT	O54943		1-1257	N-term 6xHis pDEST26
PER2 ΔCterm	O54943		1-1221	N-term 6xHis pDEST26
PER2 core PER1 CBD	O54943	1105-1209	1-1112	N-term 6xHis pDEST26
PER2 core PER1 CBD+tail	O54943	1105-1291	1-1112	N-term 6xHis pDEST26
PER2 core PER1 CBD PER2 tail	O54943	1105-1209, 1216-1257	1-1112	N-term 6xHis pDEST26
dpPER	Q7Z0C9	926-1056		N-term 6xHis-GST-3C

Table 21: Cells

Strain	Genotype	Manufacturer	Purpose
E. coli			
DH5α	F$^-$ Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17($r_K m_K^+$) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ$^-$	Novagen®	Cloning and plasmid amplification
Rosetta (DE3) pLysS	F$^-$ ompT hsdS_B(r_B^- m_B^-) gal dcm lacY1 (DE3) pLysSRARE* (Cam^R)	Thermo Fisher	Recombinant Protein expression
DH10BacY	F$^-$*mcr*A Δ(*mrr*-hsdRMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara*, *leu*)7697 *gal*U *gal*K λ$^-$ *rps*L *nup*G/pMON14272/pMON7124	Thermo Fisher	Recombinant Bacmid generation
Insect			
Sf9		Thermo Fisher	Virus production, recombinant protein expression

***** pRARE encodes the tRNA genes *argU, araW, ileX, glyT, leuW, proL, metT, thrT, tyrU* and *thrU*. tRNAs for the rare codons AGG, AGA, AUA, CUA, CC and GGA are supplemented.

3 Methods

3.1 Tree construction

The species set from Kotwica-Rolinska et al (2022)[169] was used. The dataset was extended with CLOCK and BMAL sequences retrieved from NCBI using BLAST (blastp and tblastn) against nr and transcriptome shotgun assembly (TSA) databases. Query sequences originated from the high-quality genome assemblies and respective proteomes of *Danaus plexippus* (Monarch butterfly), *Mus musculus* (mouse), *Homo sapiens* (human), *Platynereis dumerilii* (the moonstruck worm), *Danio rerio* (zebra fish), *Drosophila melanogaster* (fruit fly), *Xenopus laevis* (frog), *Pyrrhocoris apterus* (firebug), *Apis mellifera* (honey bee), *Acyrthosiphon pisum (pea aphid)*. The phylogenetic tree was build using timetree.org.[163] [CITATION of relevant papers, timetree is based on]. Timetree.org unfortunately couldn't resolve all taxa, those species were omitted. ARNT, BMAL1, BMAL2, and CLOCK were assigned in unclear cases by alignment with established homologues. These proteins have distinct C-termini and PAS domains. If assignment was unsuccessful, it is stated. If multiple partial sequences were found all were used for analysis. Sequences were aligned using CLUSTAL OMEGA by EBI[162], visualized in Jalview[159] and multiple sequence alignments were plotted using ENDscript[165]. Interaction motifs were based on literature. Interactive tree of life (iTOL) was used to plot the protein/interaction sites on the phylogeny.[164]

3.2 Protein structure prediction by AlphaFold multimer

Structures of protein complexes were predicted using AlphaFold multimer 2.3, a state of the art protein prediction tool[170,171]. AlphaFold requires only the protein sequences as input and will calculate several independent models, which are then relaxed and scored. The prediction score per model and the score per amino acid is reported. The top ranked model is shown in the figures here and a score based positional coloring is shown in the annex. Models were aligned and visualized in ChimeraX[160,161].

3.3 RNA extraction

RNA extraction was used to isolate mRNA for cloning of coding sequences of cryptochrome genes from a range of insects.

Trizol (1 mL) was added to animals or heads and the samples homogenized with a plastic pestle inside a 1.5 mL reaction tube and incubated for 5 min at room temperature. Chloroform (200 µL) was added per mL Trizol, vigorously shaken and incubated for 3 min. The samples were centrifuged (12000 x *g*, 15 min, 4°C) to separate the organic and aqueous layer. RNA is contained in the aqueous layer, while DNA and proteins are found in the red organic layer. The top aqueous layer was carefully pipetted off, RNase-free glycogen (1 µL) and isopropanol (500 µL) added and incubated for 10 min to precipitate RNA. After centrifugation (12000 x *g*, 10 min, 4°C) the RNA was found as a white pellet at the bottom

of the tube. Ethanol (75%, 1 mL) was added to the pellet and incubated for 1 h at -80°C. RNA was pelleted again by centrifugation (7500 x *g*, 5 min, 4°C), supernatant discarded and air dried for 5-10 min. RNA was dissolved in 20 µL RNase-free water. The Qiagen RNeasy Mini-Kit was used to further purify the crude total RNA isolated by Trizol. The kit's protocol was followed, and RNA eluted from the column with 30 µL RNase-free water and stored at -80°C.

3.4 cDNA synthesis

mRNA can be reverse transcribed to a complementary DNA strain (first-strand synthesis) by reverse transcriptase (RT). This is an effective way to generate DNA corresponding to expressed genes without introns ready to clone into expression vectors. Reverse transcription can be done using gene-specific primers (GSP), random hexamer primer or oligo(dT) annealing to the polyA tail of mRNAs.

The mMLV-based reverse transcriptase (Superscript II RT comparable) from the IMB Protein production core facility was used and the supplied protocol followed:

Table 22: cDNA synthesis

Step		**Oligo(dT) 12-18mer**	**Random hexamer**	**GSP**
1 Mix	primer	1 µL (500 ng)	1 µL (250 ng)	2 µL (20 pmol)
	Total RNA	1 µg		
	dNTP mix	1 µL		
	RNase-free water	To 12 µL		
2		Heat for 5 min at 65°C, then quickly chilled on ice		
3		Add 4 µL 5x first-strand buffer, 2 µL DTT (0.1M), 2 µL RNaseOut, mix		
4	incubate	42°C, 2 min	25°C, 2 min	42°C, 2 min
5	add	1 µL RT		
6	incubate	42°C, 50 min		
7	inactivate	70°C, 15 min		

Materials from Table 9.

The cDNA was run on analytical agarose gel electrophoresis and used for PCR followed by cloning.

3.5 Molecular Cloning

Gibson assembly was used for plasmid construction[172,173]. PCRs were performed under standard conditions. PCR amplicons were mixed, 2x Gibson Master mix from IMB PPCF added and incubated for 1h at 50°C for assembly. The mixture was then transformed in chemically competent *E.coli* DH5α and plated on LB-agar with the appropriate selection antibiotic.

Colonies were screened by colony PCR or in cases of a small number of colonies, picked for mini-prep.

3.6 Preparing chemically competent *E. coli*

Competent cells are able to take up plasmid DNA, necessary for cloning, plasmid production and to generate recombinant bacmid/baculovirus.

Cells were streaked on LB-agar. A single colony was picked to inoculate a preculture LB. A main culture (100 mL, LB, Table 4) was inoculated with preculture to OD600 ~ 0.1 and grown to OD600 ~0.6 at 37°C. Cells were incubated for 5 min on ice, harvested by centrifugation (4000 rpm, 10 min, 4°C) and resuspended in prechilled Tbf1 (8 mL, Table 3). Cells were centrifuged (4000 rpm, 5 min, 4°C) and resuspended in prechilled Tbf2 (5 mL, Table 3), incubated for 15 min, aliquoted (50 mL), frozen in liquid nitrogen and stored at -80°C.

3.7 Transformation of chemically competent *E. coli*

Plasmid DNA is added to chemically competent *E.coli* DH5α cells (50 µL), previously thawed on ice. Cells are incubated on ice for 15 min and heat-shocked (42°C for 45 s). 950 µL LB is added and cells are incubated for 1 h at 37°C. Cells are spun down, majority of medium discarded, cells resuspended in remaining medium and plated on a LB agar plate with appropriate antibiotics.

3.8 Mini-prep of plasmid DNA from *E. coli*

Plasmid DNA preparation was performed as described in the manual of the manufacturer. Briefly, cells are harvested by centrifugation, resuspended in defined neutral buffer containing RNase (250 µl), lysed in SDS and sodium hydroxide (250 µL) and neutralized with potassium acetate (350 µL) yielding potassium-SDS, an insoluble soap which precipitates most cell components including membrane associated genomic DNA. Precipitate is pelleted by centrifugation and supernatant loaded on silica-based columns to which the plasmid DNA binds. Column associated DNA is washed with ethanol and eluted from the column by water.

3.9 PCR

The high fidelity polymerase from IMB's protein production core facility was used to generate PCR products for Gibson Assembly.

Table 23: Pipetting scheme of high-fidelity PCR.

Reagents	**Volume (µL)/ 50 µl reaction**	**Final reaction concentration**
10x Buffer	5	1x
HF-DNA Polymerase (2 cU/µl)	0.5	1 cU
Forward primer (10 µM)	2.5	500 nM
Reverse primer (10 µM)	2.5	500 nM
dNTPs (10 mM)	1	200 µM
ddH_2O	37.5	-
DNA template (50 ng/µl)	1	1 ng/µl

Reagents are mixed besides template (Table 23) and aliquoted to the necessary number of PCR tubes. Template DNA is added. DNA is amplified in a PCR cycler using five steps:

1. Initial DNA denaturation (melting) at 98°C for 30 s
2. DNA denaturation at 98° for 30 s
3. Primer annealing at 60°C for 30 s (primer specific temperature)
4. Elongation at 72°C for 30 s/kb
5. Final elongation at 72°C for 5 min

Steps 2-4 are repeated 30 times leading to exponential amplification.

3.10 Colony PCR

Colony PCR was used to check the presence of desired DNA inserts in vector DNA (plasmid) without the need of plasmid isolation.

Table 24: Colony PCR pipetting scheme

Reagents	**Volume (µL)/ 10 µl reaction**	**Final concentration**
5x GoTaq Green buffer	2	1x
Screen-Blend DNA Polymerase mix (5 cU/µl)	0.1	0,5 cU
Forward primer (10 µM)	0.2	200 nM
Reverse primer (10 µM)	0.2	200 nM
dNTPs (10 mM)	0.2	200 µM
ddH_2O	7.3	-
Colony	1 colony	-

All components besides the colony are mixed as given in Table 24, usually in the form of a mastermix. This mastermix is aliquoted at 10 µL into PCR tubes and a single colony picked from LB-agar, dipped in the PCR mix, and streaked on a fresh LB-agar plate with appropriate selection antibiotic incl enumeration for identification. Ideally one primer anneals in a vector specific region like the tag and the other primer in a insert specific region allowing short amplification times, check for vector and check for desired insert.

PCR is amplified using a PCR cycler with the same steps as in a PCR. In the first step not only is the DNA denatured but also *E. coli* cell broken. 30-40 cycles are performed and 2 µL loaded on an analytical agarose gel.

3.11 Agarose gel electrophoresis

DNA length is checked routinely by analytical agarose gel electrophoresis. If a specific DNA fragment has to be isolated from a mixture of DNA, preparative agarose gel electrophoresis is employed.

0.8% (w/v) agarose in 1x TAE (Table 3) buffer is cooked until the agarose completely dissolves. 1:10000 SYBR Safe (e.g. 2 µL for 20 mL of gel; Table 12) is added and the gel cast in a gel chamber and a comb is added. After polymerization, the gel chamber is placed in the electrophoresis chamber, filled with 1x TAE. Comb pulled and DNA mixed with loading dye filled in the resulting pockets. A marker with DNA fragments of known size is loaded as reference. A constant voltage of 90 V and a maximal current of 150 mA is applied for 30 min. The DNA is visualized under UV light by the interaction of the SYBR Safe dye with DNA resulting in a UV fluorophore.

3.12 Protein expression in *E. coli*

Plasmid was transformed into *E. coli* Rosetta (DE3) pLysS cells. If Kan^R plasmids were used, the transformation is not plated but used to inoculate a liquid overnight preculture. If Amp^R plasmids were used the transformation has to be plated due to the only bacteriostatic nature of ampicillin and secretion of the resistance marker, β-lactamase. The plates were scraped off the next day and bacteria grown in liquid medium for 2 h to generate a preculture.

1 L of expression culture (900 mL TB, 100 mL 10x TB salts) was inoculated with typically 25 mL of preculture aiming for a OD600 ~ 0.1. Bacteria were grown at 37°C and shaking until OD600 ~ 0.6, then the temperature was reduced to 18°C. Protein expression was induced by addition of 0.25 mM IPTG. Cells were harvested after over-night expression (~20 h) by centrifugation at 5000 x *g* for 15 min at 4°C.

3.13 Blue/white screening & Bacmid DNA purification

- pCoofy27 or 28 were transformed into DH10BacY cells with an extended outgrowth step of overnight instead of 1DH10BacY cells contain (i) a transposase coding helper plasmid and (ii) a bacmid necessary for baculovirus production with a YFP reporter gene and a *lacZ* gene with internal transposition sites. The transposase catalyzed the transposition of the expression cassette from pCoofy27/28 into the bacmid resulting in a broken *lacZ* gene and therefore a lack of β-galactosidase.
- 135 µL of a 100-fold and 1000-fold dilution of the transformation was plated on a LB-agar plate containing Gentamycin, Kanamycin, Tetracycline, IPTG and X-Gal and the remaining transformation kept. Successful transposition of the expression cassette into the bacmid is seen a white colonies on a background of blue colonies with intact *lacZ* gene, yielding functional β-galactosidase which cleaves X-Gal yielding blue color after oxidation and dimerization of 5-bromo-4-chloro-hydroxyindole to 5,5'-dibromo-4,4'-dichloro-indigo.

- Two white colonies were grown overnight in LB containing Gentamycin, Kanamycin, Tetracycline and bacmid DNA was isolated by using the same buffers and volumes as for mini-prep. The supernatant after cell lysis was not applied to a column, instead isopropanol was added (700 µL) to 40% final concentration precipitating bacmid DNA. Precipitated DNA is the washed by ethanol twice (200 µl, 50 µl). Prepped bacmid DNA was then used for transfection of Sf9 cells.

3.14 Transfection into Sf9 cells and viral stock production

Sf9 cells are transfected with isolated bacmid yielding the first, low titer virus (P0). Titer is increased by passaging it (P1). Transfection, infection, and Protein expression can be followed by the YFP reporter.

Bacmids are spun again, ethanol taken off and dried in the flow hood. 20 µL water were added to dissolve the DNA, followed by 200 µL SF900 III medium. 200 µL SF900 II medium were mixed with 20 µL Cellfectin II. 100 mL of the Cellfectin dilution were added to the dissolved bacmids and incubated for 30 min.

Cell stock was diluted to 10^6 cells/mL without antibiotics. A 6-well plate was prepared with two wells for each bacmid (clone A, A', B, B'), a cell control (CC) and medium-only control (M), to have a back-up transfection in case of contamination, allowing to trace contamination, and a control for cell autofluorescence when checking transfection. All wells were filled with 2 mL of medium, but M, which got 3 mL. 1 mL of cell stock was added to A, A', B, B' and CC. 150 µL of transfection mix of clone A was added dropwise to A and A' and repeated for clone B.

Three days post-infection cells were checked for YFP fluorescence. 50 mL of Sf9 cells at 0.5-1*10^6 cells/mL were infected with 2 mL of the collected P0 virus and shaken for 3 days resulting in P1. P1 was harvested by centrifugation at 500 x *g* for 5 min. The pellet should be yellow and can be used for expression check. The supernatant was ready-to-use.

3.15 P1 expression check

Viral infection and expression were checked by thawing the P1 pellet, resuspending it in Tris-HisA (1 mL with 1% (v/v) Triton X-100) and incubating it for 10 min on ice. The lysate was centrifuged at 10000 x *g* for 10 min at 4°C. The supernatant was added to Ni-NTA beads (10 µL) in Tris-HisA and incubated for 30 min. The beads were washed three times with 1 mL Tris-HisA. Tris-HisB (25 µL) was added to elute protein. The beads were cooked with 1X SDS-PAGE loading dye (25 µL). Expression was checked by SDS-PAGE with pellet, supernatant, elution and bead samples.

3.16 Protein expression in *Sf9* insect cells

Not all proteins can be expressed in *E. coli* as they may need post-translational modifications or chaperone helping in folding. Insect cell expression is besides mammalian expression and yeast one option to circumvent such problems.

Proteins were expressed using the Bac-to-Bac baculovirus expression system using pCoofy27 and pCoofy28 backbones introducing a N-terminal His-3C tag and His-GST-3C tag, respectively[174]. 1 L of Sf9 cells at 10^6 cells/mL was infected with 10 mL of P1 virus and harvested 72 hpi at 500 x *g*, for 15 min at 4°C and cells flash frozen in liquid nitrogen.

3.17 SDS-PAGE

Sodium dodecylsulfate polyacrylamide gel electrophoresis is used to separate proteins by their molecular weight. Since folding state and intrinsic charge hinder separation by molecular weight/size, proteins are denatured by sodium dodecyl sulfate (SDS), a strong, negatively charged detergent. β-Mercaptoethanol is used to reduce potential disulfide bonds. This yields a fixed charge to mass ratio correlating electrophoretic mobility to molecular weight. To increase resolution discontinuous SDS-PAGE is used so that protein:SDS complexes are focused between the top stacking and the lower separating gel.

Table 25: Pipetting scheme for SDS-PAGE gel preparation

Substance	Stacking gel (4%) / mL	Separating gel (10%) /mL
ddH_2O	12.16	22
1 M BisTris	5.00	18
Acryl-Bisacrylamid solution	2.66	20
10% APS	0.160	0.36
TEMED	0.020	0.030l

Spacer plate and thin plate were fixed in the gel caster, separating gel components mixed, adding APS and TEMED last and mixed without introducing more air than necessary and poured in to the between the glass plates. Isopropanol (500 µL) is added to give a straight upper edge. Once polymerized isopropanol was poured off, dried and stacking gel components mixed and poured on top. Comb was pushed in and gel left to polymerize.

40 µl of protein sample were mixed with 10 µl of 5x SDS PAGE sample buffer (Table 3). The gels were placed in the SDS gel chamber which was filled with MOPS running buffer (Table 3). MW reference and samples were loaded to the gel and gel ran at 250 V for 30 min.

Gels were stained with Coomassie brilliant blue. The dye attached unspecifically to amino acids. Gels were placed in Coomassie staining solution. First signal was visible after minutes yet for permanent and full sensitivity gels were stained overnight. Gels were destained using water and a piece of tissue.

3.18 Western blot

Western Blot enables protein detection using antibody specificity and sensitivity.

Proteins were blotted from the unstained SDS-PAGE gel onto nitrocellulose and PVDF membranes as described in the "Trans-Blot® Semi Dry Instruction Manual" (Table 6, Table 8) applying 10 V for 30 min. To see if the transfer worked, bands were visualized using Ponceau S (Table 8). Blocking, washing and incubation with 5xHis-HRP antibody (Table 8) and immunodetection was executed as described in the "QIAexpress® Detection and Assay Handbook" (Table 8). The chemiluminescent detection was done, using the Molecular Imager® ChemiDOc XRS+ and Licor Imager (Table 6) as described in DETECTION REAGENT MANUAL.

3.19 Protein Purification

Proteins can be purified by many different methods where the right combination of methods has to be empirically determined for every protein. Popular methods are based on chromatography for their reproducibility especially if fast protein liquid chromatography systems (FPLC) are used. FPLCs allow method standardization and reproducibility while requiring minimal user intervention. FPLCs also allow online monitoring and recording of many parameters such as pressure, conductivity and most importantly UV absorbance, an indicator of protein concentration and purity. FPLCs also allow automated and even triggered fractionation relatable to measured parameters. Every method makes use of the protein characteristics like affinity to a ligand, charge, size and less commonly in research scale: precipitation in ammonium sulfate, solubility in organic solvents, thermal stability, hydrophobicity. The aim of every purification is a contamination-free, chemically defined protein/buffer preparation. The first step is therefore usually an enrichment of the protein by high affinity, specific interaction such as immobilized metal ion affinity chromatography or binding to a specific immobilized ligand. Tags or fusion proteins responsible for such an interaction are genetically fused to the gene of interest to yield a tagged or fusion protein. In a second step contaminants can be excluded by making use of a different characteristic of the protein such a charge or the loss of the tag/fusion responsible for binding before. In a final step the protein is separated by size in size exclusion chromatography to separate monomeric, oligomeric and aggregated protein from each other, check for homogeneity and often change into defined storage buffer.

3.19.1 Cell lysis

The frozen cells were withdrawn from the -80°C freezer, thawed and resuspended in buffer A supplemented with 1 mM PMSF and 5 mM $MgCl_2$. In case of *E. coli* expression, 2 µL of Sm nuclease/100 mL suspension and 10 mg lysozyme were used and in case of insect cells 1 protease inhibitor cocktail tablet and 1 µL Sm nuclease/100 mL suspension were added. The suspension was

stirred until homogenous and put on ice. Cells were lysed by sonication. Insect cells only have a membrane and therefore lyse much faster (2-3x 1 min, output 6, duty 50%) than *E. coli* with their cell wall (5x 1 min, output 6, duty 50%). Lysates were cleared by centrifugation (50000 x *g*, 30 min, 4°C).

3.19.2 Immobilized metal ion affinity chromatography (IMAC)

Immobilized metal ion affinity chromatography (IMAC) is probably the most widely used method in research labs for the first purification step. IMAC is based on the complexometric interaction of amino acids (usually hexa-histidine) with immobilized metal ions (usually Ni^{2+} or Co^{2+}) which are immobilized in a matrix of usually Nitrilotriacetic acid (NTA). NTA only takes up four out of six coordination sites which will be occupied by the hexa-histidine tag (bidentate ligand). The interaction is very specific and has a high affinity and is therefore ideally suited as the first purification step. The hexa-histidine tag is very small and usually doesn't interfere with folding and function of the protein and can easily be removed in combination with a specific protease cleavage site. Competitive elution with a cheap monodentate ligand, imidazole, is used.

Buffer B was added to 10% yielding 30mM imidazole important to suppress unspecific binding to the column. The lysate is applied to the equilibrated column at 2.5 mL/min (2 min contact time), then the column is washed with buffer A and in a second wash step with 10% buffer B. The protein can be eluted in a step to 50 or 100% buffer B or a continuous gradient, typically 20 column volumes to 100% buffer B. Step elution was employed when protein was known to be very pure after IMAC or when time and small volume were important knowing that an ion exchange chromatography would yield sufficiently pure protein.

3.19.3 GST Affinity chromatography

Like IMAC purification by GST fusion is an affinity method. Glutathione-S-Transferase is a protein that naturally binds glutathione (GSH), a cellular antioxidant and tripeptide (γ-glutamylcysteinylglycine). Genetic fusion of the GST gene to the gene of interest yields a GST-fusion protein which binds matrix bound GSH. This very specific interaction yields very pure protein since E. coli doesn't have a GST protein. GST usually also increases solubility and expression yield.

Lysate was applied to column at ~2 mL/min. Since the GSH column is self-packed (XK16, 60 mL) it cannot be connected to a FPLC system, and the flowrates are estimates. The column was washed until no lysate was washing off and 50 mL of GST-wash was applied to reduce chaperone, DNA and other protein contaminants. The column was washed with 3-5 column volumes (CV) of GST-lysis. A His-3C protease dilution was applied and circulated overnight over the column at decreased flow rate. The next day the circulate was washed off containing the protein of interest by applying 1.5 column volumes (90 mL). Bound GST (and uncut fusion protein) was eluted with 1.5 CV (90 mL) GST-E buffer.

3.19.4 Ion-exchange chromatography

Ion exchange chromatography (IEX) is based on binding of the protein in a low salt condition to a charged matrix and eluting it in an increasing salt gradient. Cation exchange (immobilized carboxyls) and anion exchange (immobilized amines) matrixes are available and chosen based on pH, theoretical pI and if necessary, empirically.

IEX columns were equilibrated with water, 100% buffer B followed by 0% buffer B (100% buffer A) until the conductivity was stable. IMAC eluates were diluted 5-10 fold in buffer A to reduce the salt concentration enough to bind to the column, usually <50 mM NaCl or 5 mS/cm or as determined empirically. CRYs tend to precipitate in low salt and were therefore diluted step wise and applied at a high flow rate (5 mL/min). Column was washed with five column volumes (CV) of buffer A or until conductivity was stable. The column was developed with a gradient of 0-50% B over 10-20 CV. Column was washed with 100% B and equilibrated with 0% buffer B, ready for the next IEX.

3.19.5 Size exclusion chromatography

Size exclusion chromatography separates molecules based on their hydrodynamic radius as they pass from an inert porous matrix of crosslinked agarose. Larger molecules cannot enter the pores as much as smaller molecules resulting in a shorter path and earlier elution. Smaller molecules enter the pores and therefore have a longer retention time/elute later. Smaller sample volumes increase resolution and typically 2-4% of column volume are loaded.

Concentrated proteins were centrifuged (10000 x *g*, 10 min, 4°C) and applied to the equilibrated sample loop over the sample inlet using a syringe (injection valve position: load). The equilibrated column was pre-run until the pressure was stable before applying the sample by running buffer over the sample loop (injection valve position: inject). Once double the loop volume ran over it, injection valve position is set back to load and the column was developed by one column volume of buffer.

Table 26: Purification steps for all proteins used. Used buffers and concentrator molecular weight cut off (MWCO) per protein. Buffers refer to Table 5.

Step/Buffer	Mouse CRY	Insect CRY	Monarch CRY2	Monarch CRY2:PER	Mouse PER1	Mouse PER2
N-terminal tag	PHR: His_7-3C Fl: none	His_7-3C	His_7-GST-3C	His_7-3C His_6-GST-3C	His_6-GST-3C	His_6-3C
Cleaved?	no	no	yes	yes	yes	no
C-terminal tag	PHR: none Fl: His_6	none	none	none	none	none
IMAC						
Buffer A	HisA-Tris	HisA-Tris	HisA-SPG			HisA-Tris
Buffer B	HisB-Tris	HisB-Tris	HisB-SPG			HisB-Tris
step/grad?	gradient	step	gradient			gradient
GST						
Buffer A				HisA-SPG	GST-lysis	
Buffer wash				GST-SPG-wash	GST-wash	
Buffer E				GST-SPG-E	GST -E	
IEX	SP HP		SP HP			Q HP
Diluted…	10x		10x			10x
Buffer A	IEX-A-Tris		IEX-A-SPG			IEX-A-Tris
Buffer B	IEX-B-Tris		IEX-B-SPG			IEX-B-Tris
SEC	SEC	SEC	SEC-SPG	SEC-SPG	SEC	SEC
MWCO	30	30	30	10	10	10

3.20 DNA and protein quantification

Protein and DNA concentrations were determined spectroscopically. DNA has an absorbance maximum at 260 nm commonly used for quantification. Many methods exist to quantify proteins based on interactions of amino acids with dyes (e.g. Bradford) or reduction of metals (bicinchoninic assay, BCA) but rely on correlating signal to known amounts of a standard protein. Since proteins differ in composition and therefore reactivity to such reagents, the spectroscopic method is superior taking composition into account. Tryptophan, tyrosine and too much lesser extent phenylalanine and histidine all absorb ultraviolet light at 280 nm. Tryptophan has a peak at 280 nm and tyrosine at 276 nm. From the known composition of the protein an extinction co-efficient can be calculated using software such as ProtParam or SnapGene (Table 15 and Table 16). The two programs produced almost the same result. For consistency the values calculated by SnapGene were used. Absorbance A, extinction co-efficient ε and concentration c are connected over the Lambert-Beer law with the known value range:

$$\mathrm{A} = \mathrm{c} \times \mathrm{d} \times \varepsilon$$

Where A is defined as $\log_{10} = \frac{I_0}{I_1}$, log(incident intensity/transmitted intensity).

The nano spectrometer (Table 6) was used to measure absorbance spectra and the values at 260 nm and 280 nm were used to calculate DNA and protein concentrations, respectively. The ratio of 260/280 was used additionally to assess protein purity where a value of 0.55 is considered very good and 0.6 as good. The spectrometer is first blanked with the respective buffer and measurements are performed in triplicate after clearing protein samples from aggregates by centrifugation (10000 x *g*, 10 min, 4°C) and applying 2-3 µL/measurement.

3.21 Isothermal Titration Calorimetry (ITC)

Isothermal titration Calorimetry measures the resulting temperature difference in a binding experiment (titration of one ligand to receptor) enabling to determine dissociation constant K, enthalpy ΔH and by calculation free Gibbs enthalpy ΔG and entropy ΔS.

$$\Delta G = \Delta H - T\Delta S$$

$$\Delta G = -RTln(K)$$

Where R is the gas constant and T temperature in Kelvin

Proteins were dialyzed against the same batch of SEC buffer for >16 h to ensure matching buffers (Table 5). Protein concentrations were determined, and proteins diluted if necessary. Proteins were pipetted in their respective well in the 96-well plate and cooled at 4°C until measured. PER served as titrant (19 injections of 2 µl every 150 s) into cryptochrome. Experiment were conducted at 25°C and performed in triplicates. Analysis was carried out using the Microcal PEAQ-ITC Analysis software. Data plotted using GraphPad Prism.

3.22 Fluorescence Polarization

50nM of FAM labeled peptides were mixed with serial dilutions of protein in 20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 0.05% Tween-20. Fluorescence polarization was measured in a Tecan Spark 20M plate reader at excitation 485/20 nm, emission 520/10 nm, gain optimal, dichroic mirror 510 nm, 30 flashes, 40 µs integration time, Z-position set to 20000 µm, at 20°C.

Data was fitted using GraphPad 9.5.0 “One Site –Total binding” model:

$$\mathrm{Y} = \frac{\mathrm{B_{max}} \cdot \mathrm{c}}{\mathrm{K_D} \cdot \mathrm{c}} + \mathrm{NS} \cdot \mathrm{c} + \mathrm{background}$$

Where:

B_{max} is the estimated maximum binding in the same unit as Y,
c is titrand concentration,
K_D is the dissociation constant, the concentration of half-maximum binding,
NS is the slope on nonspecific binding,

Background is a constant term. The free fluorescent probe yields a non-zero signal.
Y is the signal in millipolarization (mP) as calculated by the Tecan Spark Control software.

$$Y = P = \frac{F|| - F\perp}{F|| + F\perp}$$

50 nM probe were chosen based on available literature[62,83,96,175–177], to yield 10-100-fold signal over buffer and well background and to keep the probe concentration as low as possible. The model is only valid if the titrant concentration is virtually identical to free titrant because of low probe concentration (here 50 nM). Data was plotted using GraphPad 9.5.0.

3.23 Crosslinking

Bis(sulfosuccinimidyl)suberate or short BS3 is a traditional homobifunctional crosslinker with amine reactivity on both ends of the molecule (11.4 Å) realized by sulfo modified NHS esters for water solubility.

A 10mM BS3 stock solution was prepared with SEC buffer (Table 5), importantly free of primary amines. The CRY:PER complex at 11 µM was incubated as recommended by the manufacturer at 5x, 10x, 25, and 50x molar excess of BS3 over protein complex and incubated for 2 h on ice. The NHS esters were quenched by addition of Tris to 50 mM and incubation for 15 min at room temperature. . Samples were mixed with 5x SDS sample buffer and resolved by SDS-PAGE and Coomassie stained. Mass-shifted bands were excised from the gel and submitted to Dr. Jiaxuan “Jimmy” Chen.

3.24 Photo-Crosslinking

Besides the classical homobifunctional crosslinker BS3 a heterobifunctional photo-crosslinker was used combining amine-reactive NHS chemistry and non-restricted photo-reactivity by a diazirine. Diazirines lose N_2 upon UV radiation resulting in a reactive carbene which can react with every amino acid by integrating into C-H, O-H or N-H bonds. The reactive carbene is short-lived reacting with water if no target molecule is available yielding the unique non-restricted yet specific character of photo-crosslinkers.

Here, Sulfo-NHS-Diazirine or sulfosuccinimidyl 4,4'-azipentanoate or short Sulfo-SDA was used (Table 7), where the sulfo group increases solubility in aqueous solution. A 10mM stock solution was prepared with SEC buffer (Table 5), importantly free of primary amines. The CRY:PER complex was used at 11 µM. The manufacturers protocol recommends to use 25-50x molar excess for protein complexes

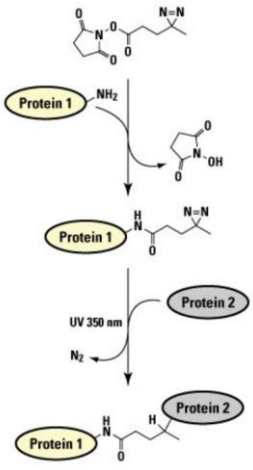

Figure 12: Sulfo-SDA reaction scheme. from product page

<5 mg/mL which was the case here. 0x, 5x, 10x, 20x and 50x molar excess of sulfo-SDA:complex was tested. Sulfo-SDA and complex where mixed and incubated for 30 min at room temperature (amine reactivity), followed by a quenching of the NHS-ester by addition of Tris to 50 mM and incubation for 15 min on ice. Photo-crosslinking was performed by irradiation of the opened tubes in a prechilled metal block with 3.6 j/cm^2 at 365 nm. The irradiation dose was tested to be sufficient. Samples were mixed with 5x SDS sample buffer and resolved by SDS-PAGE and Coomassie stained. Mass-shifted bands were excised from the gel and submitted to Dr. Jiaxuan "Jimmy" Chen at IMB Proteomics core facility for analysis by mass-spectrometry.

3.25 Photoaffinity labeling

BMAL TAD peptides were N-terminally modified with 4-[3-(Trifluoromethyl)-3H-diazirin-3-yl] benzoic acid activated by NHS by Dr. Fabian Barthels from the Pharmacy, Johannes Gutenberg University Mainz. 4-[3-(Trifluoromethyl)-3H-diazirin-3-yl] benzoic acid generate carbene units by UV irradiation <360 nm and are thought to be inactivated by water when target molecules are absent therefore lacking non-specific crosslinking.[178,179]

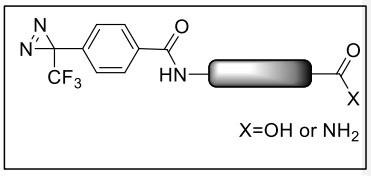

Figure 13: Scheme of modified peptide. Image courtesy by Dr. Fabian Barthels.

Purified protein (complex) was incubated with a 3x fold molar excess of modified peptide to saturate the protein in 1.5 mL reaction tubes for 15 min on ice. Tubes were opened, placed in a prechilled metal block ensuring upright orientation of the opening and irradiated with 3.6 J/cm^2 of UV light (365 nm, Table 6). The samples were mixed with 5x SDS sample buffer (Table 3) and separated by SDS-PAGE while keeping the 10 kDa marker band in the gel.

Mass-shifted bands were cut and submitted to Dr. Jiaxuan "Jimmy" Chen at IMB Proteomics core facility for analysis by mass-spectrometry.

3.26 Mass-spectrometry and data analysis (Collaboration)

Dr. Jiaxuan Chen performed the following steps:

3.26.1 In-gel digestion

Excised gel piece was cut into small cubes, followed by destaining in 50% ethanol/25 mM ammonium bicarbonate. The proteins were then reduced in 10 mM DTT at 56°C and alkylated by 50 mM iodoacetamide in the dark at room temperature. Afterwards, proteins were digested by trypsin in 50 mM ammonium bicarbonate at 37°C overnight. Following peptide extraction sequentially using 30%

and 100% acetonitrile, the sample volume was reduced in a centrifugal evaporator to remove residual acetonitrile. The resultant peptide solution was purified by solid phase extraction in C_{18} StageTips.[180]

3.26.2 Liquid chromatography tandem mass spectrometry

Peptides were analyzed using an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) coupled to EASY-nLC 1200 UHPLC system (Thermo Fisher Scientific). Peptides were separated in an in-house packed 55-cm analytical column (inner diameter: 75 µm; ReproSil-Pur 120 C18-AQ 1.9-µm silica particles, Dr. Maisch GmbH) by online reversed phase chromatography through a 90-min gradient of 2.4-33.6% acetonitrile with 0.1% formic acid at a nanoflow rate of 250 nl/min. The eluted peptides were sprayed directly by electrospray ionization into the mass spectrometer. Each sample was injected twice and measured using two different combinations of collision energies in stepped mode (PMID: 30693776). Mass spectrometry measurement was conducted in data-dependent acquisition mode. For BS3- and SDA-crosslinked samples, a top15 method was used with one full scan (resolution: 60,000, scan range: 300-1650 m/z, target value: 3×10^6, maximum injection time: 40 ms) followed by 15 fragmentation scans via higher energy collision dissociation (HCD; normalized collision energy in stepped mode: 25, 30, 35% or 27, 30, 33%; resolution: 15,000, target value: 1×10^5, maximum injection time: 40 ms, isolation window: 1.4 m/z). For TPD-crosslinked samples, a top10 method was used with one full scan (resolution: 60,000, scan range: 300-1650 m/z, target value: 3×10^6, maximum injection time: 60 ms) followed by 10 fragmentation scans via higher energy collision dissociation (HCD; normalized collision energy in stepped mode: 25, 30, 35% or 27, 30, 33%; resolution: 30,000, target value: 1×10^5, maximum injection time: 60 ms, isolation window: 1.4 m/z). Only precursor ions of +3 to +8 charge state were selected for fragmentation scans. Additionally, precursor ions already isolated for fragmentation were dynamically excluded for 25 s.

3.26.3 Mass spectrometry data analysis

Raw data files were pre-processed by MaxQuant software package (version 1.6.5.0)[181] as described [182]. The peak lists (*.HCD.FTMS.sil0.apl files) were searched using xiSEARCH (version 1.7.6.7)[183]. The following settings were used: enzyme specificity, trypsin/P (for BS3 or SDA) or trypsin/P+GluC (for TPD); allowed maximum number of missed cleavages (4); BS3 specificity, linking K, S, T, Y and protein N-terminus; SDA specificity, linking any amino acid to K, S, T, Y and protein N-terminus; TPD specificity, linking any amino acid to protein N-terminus; fixed modification, carbamidomethyl (C); variable modifications, oxidation (M) and mono-links for linear peptides with dead-ends; MS1 tolerance, 6 ppm; MS2 tolerance, 20 ppm; boosting option activated for residue pairs; residue-level FDR was set at 5%. For TPD-crosslinks, minimum peptide length was set to 4.

4 Results

4.1 The CRY-m:BMAL1 TAD interaction

The BMAL1 C-terminus was identified as a transactivation domain (TAD) with a minor contribution from the preceding region (G-domain, the sequence between PAS-B domain and TAD)[71]. The paralogs CRY1/2 repress BMAL:CLOCK activity by binding to the BMAL1 transactivation domain, displacing the histone acetyl transferase CBP/p300 KIX domain in the process[71,83,149,184–190] (Figure 10). The CRY1 coiled-coil (CC) helix and tail were identified as the binding site of the BMAL1 TAD by peptide scan analysis and affinity measurements[83]. The interaction of the BMAL1 TAD and CRY1 CC helix was later verified by NMR and mutational studies, which additionally suggested a negative correlation of affinity and period length[71]. Yet both studies worked primarily with peptides (of the CC-helix). Crystal structures show that the CC-helix is partially surface exposed but not sticking into the solvent (Figure 7). Later, it was shown that CRY1/2 PHR in complex with PER2 have an insignificantly different affinity to BMAL1 TAD as apo CRY1/2 PHR[96]. A surprising finding since PER2 wraps around the CC-helix[98,99] and should therefore block the BMAL1 TAD binding site at least partially (Figure 7). Here the CRY1/2 primary pocket is proposed as an additional BMAL1 TAD binding site.

4.1.1 AlphaFold

An AlphaFold multimer prediction was used to get a first idea where the BMAL1 TAD might bind. The input sequences were mouse $BMAL1_{600-632}$ TAD and $CRY1_{1-496}$ PHR (Figure 14 A). AlphaFold predicts the BMAL TAD to bind to the CRY1 primary pocket so that the N-terminus would stick out and the C-terminus would fold back onto the α-helix as shown in Figure 14 D. The prediction score (confidence) is quite low as shown in Figure 39. The model fits to the NMR chemical shifts from Xu et al (2015)[71] showing that a C-terminally attached quencher (Mn^{2+}-EDTA) would result in reduced chemical shifts for α-helix residues indicating spatial proximity of the two.

AlphaFold multimer can have a tendency to place small ligands in pockets especially if those pockets are known to be occupied in crystal structures (the training set of AlphaFold) based on personal experience and [191]. Model confidence can be increased if the same result is obtained when varying inputs are provided for example in the form of inputs of varying length containing the motif under investigation. Longer inputs for example shouldn't be folded up to fit into random pockets, because this would yield a suboptimal result in the energy calculation.

The predictions were repeated using $CRY1_{1-496}$ and the BMAL1 G domain and TAD as input with increasing the length stepwise from the TAD/C-terminus towards the C-terminus of the PAS-B domain. A second set of BMAL inputs started from the middle and expanding towards PAS-B C-terminus and TAD. The third set of inputs started from the PAS-B C-terminus and increased in length towards the

TAD. This resulted in 30 BMAL1 fragments of varying length but importantly every amino acid was sampled as often as any other as shown in Figure 14 C. The top predicted structures were taken and amino acids from $CRY1_{1\text{-}496}$ and BMAL1 in contact (< 5 Å) determined for each prediction (n=30) resulting in a list of unique pairs from each prediction. The fraction of common contacts (FCC) was calculated for each unique residue $pair_{i,j}$. Unique pair occurrences were summed up and divided by maximum occurrence of each amino acid, the sampling number (n=11).

$$FCC = \frac{\sum unique\ pair_{i,j}}{11}$$

The fraction of common contacts which was plotted against the BMAL1 and CRY1 amino acid position are shown in Figure 14 A. PyMOL and Python scripts are shown in Script 1, Script 2, Script 3. This approach is based on a benchmarking study by Chop Lee using the ELM database[191]. AlphaFold multimer continues to predict binding of the BMAL1 TAD to the CRY1 primary pocket as shown in Figure 14 A+D+G. Additionally, it also predicts an interaction of the BMAL1 G-domain α-helix to bind to the end of the CRY1 CC-helix in a similar way as PER2 as shown in Figure 14 G (compare to Figure 7). The individual prediction scores are low but the recurring placement of the BMAL1 TAD in the pocket increases certainty. While the conformations of CRY1 and BMAL1 TAD are almost identical, the relative orientation within the pocket is diverse. 25 predictions per BMAL1 fragment were performed, all of which show slightly different BMAL1 conformations. The conformations also change between BMAL1 input fragments as shown in Figure 14 D. This can be interpreted as multiple stable conformations within a flexible ensemble, or snapshots from a dynamic interaction. AlphaFold, just like every other prediction tool (and x-ray crystallography), cannot show dynamics but only stationary pictures. Different conformations therefore indicate flexibility in the binding. The primary pocket is larger than the BMAL1 TAD allowing multiple conformations. The interaction is mostly based on hydrophobic and electrostatic interactions. The primary pocket is lined with positively charged amino acids and the TAD is negatively charged. Multiple conformations can fulfill the charge pattern. Results are plotted on a CRY-m multiple sequence alignment (MSA) in Figure 17.

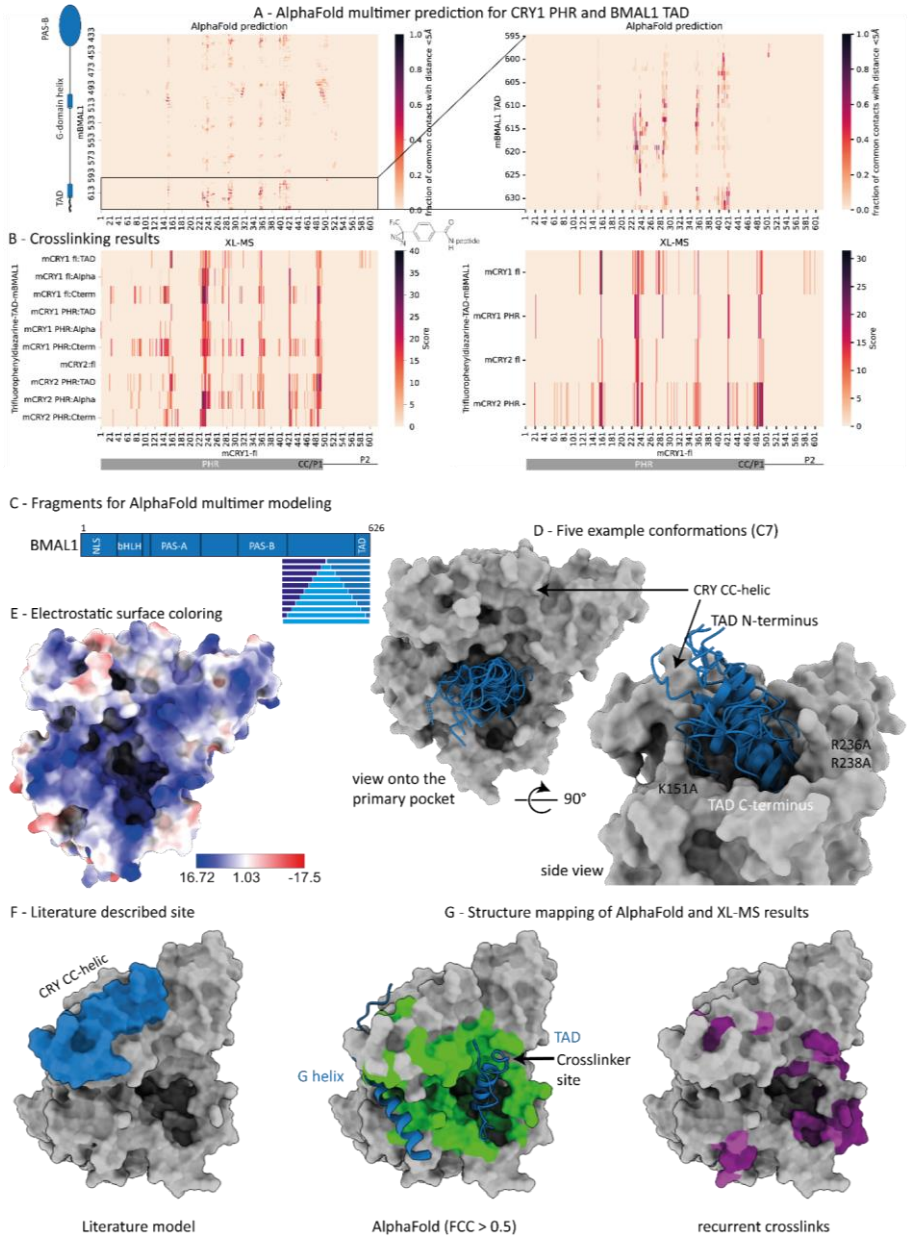

Figure 14: Results of CRY1:BMAL1 AlphaFold multimer prediction and XL-MS. A: Heatmap of fraction of common contacts (FCC) plotted against CRY1 residues (x-axis) and BMAL1 residues (y-axis). A zoomed in version focusing on the BMAL1 TAD is shown on the right. B: Heatmap of the photoaffinity labeling results between CRY1/2 fl/PHR and TPD-BMAL1 TAD and fragments with CRY1 residues on the x-axis. CRY2 crosslinks are depicted on the aligned position in CRY1. The more reliable TAD crosslinking is

shown on the right. C: BMAL1 domain scheme and fragments used as input for the AlphaFold predictions. D: example of BMAL1 TAD conformations inside the CRY1 primary pocket. Shown are the top five models of the BMAl1 TAD (blue) as cartoon from fragment C7 (TAD plus a few amino acids) with CRY1 from the best model in surface representation (grey). E: electrostatic surface coloring of the CRY1 surface showing the positive charge of the primary pocket. F: The described binding site of BMAL1 TAD on CRY1, the CC-helix, is highlighted in blue (Czarna, Breitkreuz et al (2011)[83], Xu, Gustafson et al (2015)[71]). G: Results of AlphaFold predictions and PAL are shown on the CRY surface. Left: Only positions predicted to be in contact with the BMAL1 TAD with a FCC > 0.5 are shown in green. Right: TPD-BMAL1 TAD crosslinks are shown in purple for crosslinks found in at least 50% of paralogs PHR/fl, respectively.

4.1.2 Photoaffinity labeling

Binding was measured by photoaffinity labeling (PAL). In PAL the ligand is modified with a photo-crosslinker moiety, here a diazirine was used. Upon UV irradiation diazirines lose nitrogen (N_2) resulting in a short-lived carbene (diradical) which inserts into C/N/O-H bonds. The short-lived carbene yields non-restricted (by amino acid side chain) but binding site-specific crosslinking (photo crosslinking)[178,179]. N-terminally trifluoro methyl phenyl diazirine (TPD) modified BMAL1 TAD was used. Additionally, the BMAL1 binding epitopes found by Xu, Gustafson et al (2015)[71] were tested, the α-helical part of the TAD and the C-terminus as shown in Table 5 and Figure 16 A. Binding to CRY1 and CRY2 full-length and PHR were tested and crosslinks were identified by mass-spectrometry. Mass spectrometry and peptide identification were done in collaboration with Jiaxuan Jimmy Chen from the IMB Proteomics core facility. The results are shown in Figure 14 B, G as a heatmap plotted against position in the protein and as colored surface on a predicted structure. An AlphaFold result was used to use the same CRY1 model for visualization of AlphaFold models and PAL.

BMAL1 TAD crosslinks are more specific than crosslinks of the BMAL1 TAD α-helix or the BMAL1 TAD C-terminus. The TPD moiety is located at the N-terminus and therefore the same position in the BMAL1 TAD and TAD α-helix peptides. The TAD α-helix crosslinked positions are similar but broadened compared to the TPD-BMAL1 TAD. The TPD moiety in the BMAL1 TAD C-terminus peptide is located where the BMAL1 TAD hinge region would be located according to the predictions. Those crosslinks are the least specific. This shows how fragments of a binding motif can lose their binding specificity. Further analysis is therefore based on the TPD-BMAL1 TAD crosslinks shown in Figure 14 B right.

The TAD crosslinks are located around the primary pocket with one crosslink being in the pocket at H355. In the CRY1 fl sample crosslinks to the CRY1 tail are found which interestingly overlap with the P2 peptide identified by Czarna, Breitkreuz et al (2011)[83] as a BMAL binding epitope by peptide scan analysis. We also find crosslinks on the CC-helix which is located close to the primary pocket. Results are plotted on a CRY-m multiple sequence alignment (MSA) in Figure 17.

4.1.3 Mutants

To further validate the binding site, a mutagenesis study was performed. CRY1 positions were mutated that according to the AlphaFold prediction and crosslinking result are in contact with the BMAL1 TAD. A focus lied on positively charged amino acids because of the suspected electrostatic steering mechanism of the interaction. Residues were mutated to alanine, because larger or negatively charged residues could have been detrimental to the protein fold. The affinity of these CRY1 mutants to a fluorescently labeled BMAL1 TAD (FAM-BMAL1 TAD) were tested by fluorescence polarization (FP).

Repressive activity of these mutant CRY1 proteins towards BMAL1:CLOCK was tested in human HEK293T cells using a reporter gene assay as part of a collaboration with the Kramer Lab (Charité, Berlin) and were performed by Astrid Grudziecki.

4.1.3.1 Reporter gene assay

CLOCK:BMAL1 activate transcription of luciferase under control of an E-box promoter. Luciferase activity is measured by luminescence after substrate addition and used as a proxy for transcription activation. CRY1 represses CLOCK:BMAL1 activity resulting in reduced luminescence. Here it was important to not use too much CRY1 (in the form of transfected plasmid) since these experiments follow a typical sigmoidal dose-response curve similar to what is shown in Figure 15 B and the expected effect size is rather small. Too much CRY1 would lead to the floor effect, too strong repression to see lifted repression by reduced CRY1 binding. Too less CRY1 would not yield enough signal to see reduced repression. The aim is to get around 50% repression with CRY1 WT.

Strong effects were not expected from the single mutants because probably many amino acids are important for binding and rather their charge not their identities or positions are important for binding. Results are shown in Figure 15 A. No effect is observed for CRY1 K151A and K159A compared to CRY1 WT. K159A is relatively far away from the primary pocket. CRY1 phosphate binding loop mutants R227A+K228A and R236A+R238A show higher relative reporter activity indicating reduced repression. H355A (primary pocket) , Q407E+Q408E, H411A, and R421A (C-terminal lid) further reduce reporter activity compared to CRY1 WT indicating enhanced repression. R421A seems to be dominant over R236A+R238A in the triple mutant. While all these mutants were expected to elevate reporter activity by reducing CRY1:BMAL1 affinity and thereby reducing transcription activation, only R227A+K228A and R236A+R238A achieved this.

4.1.3.2 Affinity measurement

$CRY1_{1\text{-}496}$ PHR plasmids were mutagenized, virus produced, proteins expressed in insect cells and purified with the same protocol as shown in Table 26. To increase reproducibility, the same purification protocol and batch of SEC buffer were used. Two batches of IMAC and CatEx buffers were used. 3-4

proteins were expressed and purified in parallel for standardization. Flow paths and columns were washed extensively between protein mutants to avoid cross contamination. The extinction coefficient for all these mutants was identical because no aromatic amino acid was mutated. Proteins were concentrated to similar concentration as possible by yield and assay requirement. Proteins were >95% pure after size-exclusion chromatography (SEC).

CRY1:BMAL1 TAD affinity is reduced 3.8-5.5 fold in CRY1 R236A+R238A, H355A and Q407E+Q408E. CRY1 R227A+K228A was not tested because the protein couldn't be expressed even after multiple attempts. The triple mutant CRY1 R236A+R238A+R421A wasn't cloned. Cloning for the reporter gene assays was a bit faster than cloning and virus production for the affinity measurements. Therefore, CRY1 R236A+R238A+R421A was included in the reporter gene assay. Interestingly, the polarization at saturation is not the same for all mutants. The polarization is dependent on the fluorophore and size of the fluorophore:protein complex[176]. The same fluorophore and protein size was used. Single and double amino acid exchanges do not change protein size that much for monomeric proteins of this size. SEC didn't suggest a change in oligomeric state for the mutants

Three of seven expressing CRY1 mutants show a negative effect in affinity measurements. The other four are neutral. Only one these three, CRY1 R236A+R238A (phosphate binding loop), overlaps with the CRY1 mutants that reduce CRY1-mediated repression of BMAL1:CLOCK activity. In the FP assay, protein concentrations, components, and the composition of buffer and mixture are known. It is a very reductionist and highly controlled environment with a minimalistic, chemically defined BMAL1 TAD peptide instead of a big protein. In the cellular assay full length proteins are used, protein stability cannot be controlled and most importantly many other proteins are present. The CRY1 primary pocket is not only a potential binding site of the BMAL1 TAD but also a known binding site of FBXL3, the E3 ubiquitin ligase responsible for CRY-m degradation[42,107]. Mutations around the primary pocket are known to perturb FBXL3 binding[42,98].

Reduction in FBXL3 binding by mutations in the binding site could result in decreased degradation and therefore elevated CRY1 activity in the cell. Different protein concentrations would make comparisons between mutants difficult. Indeed the positions H355, Q407+Q408, H411 and R421 are interacting with FBXL3 and have a repression enhancing effect (Figure 15 D, Table 27). H355 orders the FBXL3 C-terminus inside the primary pocket. Q407 forms a water bridged hydrogen bond to the FBXL3 backbone at M424. R421 forms a salt bridge to FBXL3 E368. No hydrogen bonding is found between CRY-m H411 and FBXL3 but it sits right under the β-sheet. Substitution with alanine would likely result in a packing defect. The protein levels in the reporter gene assay were reported to be too low for reliable quantification. Importantly, the CRY1 R227A+K228A protein was expressed in the reporter gene assay according to Western blot. Mutants are plotted on a CRY-m MSA in Figure 17.

According to Schmalen et al (2014)[98] none of the mutated residues is directly in contact with PER2.

Table 27: Results of the mutant study. Successful expression insect cells (+), measured affinity of the interaction and the result of the reporter gene assay are summarized. N. d. not determined, n. a. not applicable. Affinity was measured three times, the result of one representative measurement is given.

CRY1 mutant	**Expression in insect cells**	**K_D /FP mean (SEM)/nM**	**Repression compared to WT**	**Affects FBXL3**
WT	+	123 (6)	n. a.	n. a.
K151A	+	143 (11)	0	Not in contact
K159A	+	150 (13)	0	Distant
R227A+K228A	-	n. d.	-	Distant
R236A+R238A	+	682 (55)	-	Distant
H355A	+	469 (42)	+	W428
Q407E+Q408E	+	537 (63)	+	M424 amide binding lost
H411A	+	84 (5)	+	packing
R421A	+	117 (6)	+	E368 salt bridge lost

4.1.4 Evolutionary conservation of interaction

The circadian clock is very conserved in animals, so I wondered how conserved is this interaction?

4.1.4.1 Sequence conservation of BMAL TAD

BMAL is surprisingly well conserved at its C-terminus considering that the preceding domain (sometimes called G domain[71]) does not show much conservation (Figure 16 A). The C-terminus doesn't have a recognizable tertiary structure as far as predictions and NMR data[71] go. Conservation implies selection pressure and if this is not coming from the need to keep a domain fold, then it is likely for function.

Two striking differences are observed when comparing BMAL TAD sequences of mammals to other animals (Figure 16 A), e.g. insects: (i) insects have an additional proline residue at the C-terminus and (ii) and aspartate (D) to glycine (G) exchange at the C-terminal binding epitope. If butterflies and moth (Lepidoptera) are included two more amino acid exchanges are found, again aspartate (D) to glycine (G) at the TAD N-terminus and aspartate (D) to asparagine (N) at the N-terminal end of the C-terminal epitope. *Danaus plexippus*, the monarch butterfly, has a quite unique exchange of phenylalanine (F) to isoleucine (I). D to G and F to I exchanges hint towards functional differences because those are not conservative exchanges (negative to uncharged and small; large, aromatic to small aliphatic). A D to N exchange keeps geometry but eliminates a negative charge, functional relevance depends on charge importance. The AlphaFold predictions above suggested the phenylalanine to play a role in packing of

the C-terminus against the α-helix. The predictions also placed the negatively charged BMAL1 TAD in the positively charged CRY1 primary pocket. The loss of negative charges could therefore result in reduced binding of insect BMAL TADs if no compensatory exchanges in their binding pocket occurred. The different BMAL TAD versions were tested against mouse and insect CRYs. The two binding epitopes, α-helix and C-terminus were also tested to see which is responsible for affinity differences as shown in Figure 16 C.

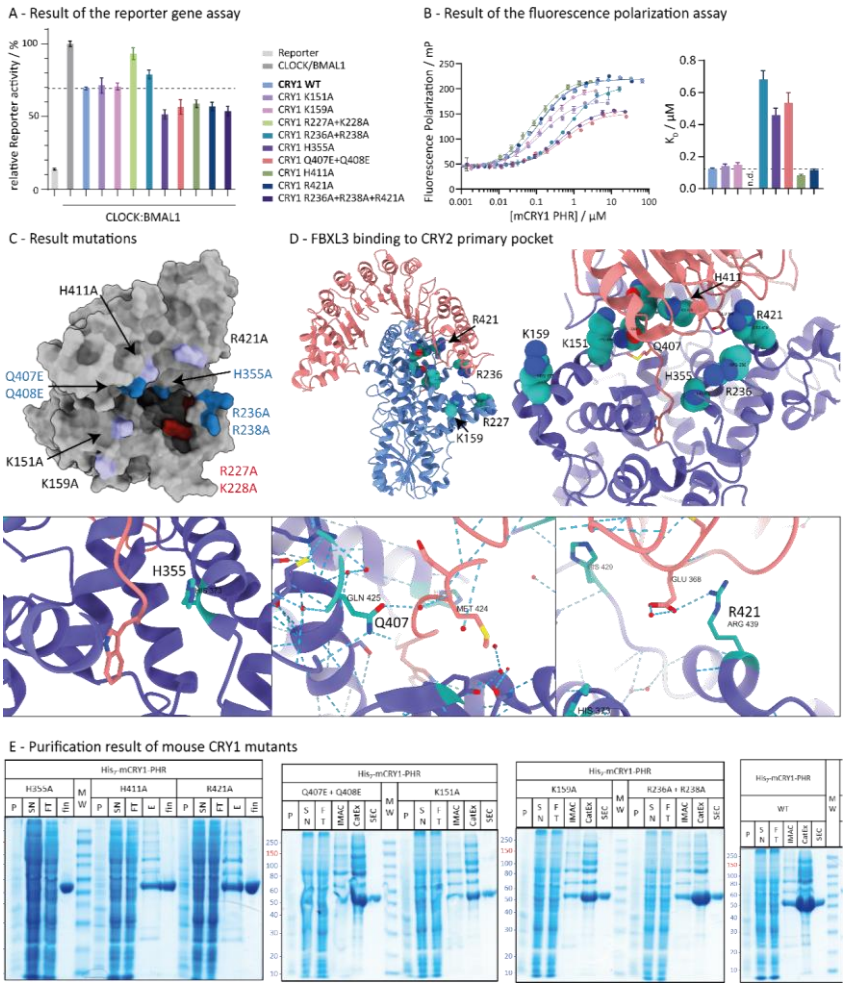

Figure 15: Results of the CRY1 mutants. A: Result of the reporter gene assay. CLOCK:BMAL activation is set to 100% and the results reported in relation. Only two CRY1 mutants show reduced repression. (n=3, mean and SEM are depicted). B: Results of the CRY1:BMAL1 TAD affinity measurements by FP. One

mutant could not be expressed. Three mutants show 3.8 to 5.5-fold reduced binding affinity (n=3, mean and SEM). C: Location of the mutations mapped on a surface representation of CRY1 with a view into the primary pocket. Dark blue mutants showed reduced binding affinity, red was not expressed in insect cells, and the light blue mutants behaved neutral. D: Interaction of CRY2 with FBXL3 (PDB 4I6J) from Xing et al (2013)[42] . The FBXL3 (light coral) C-terminus inserts deep into the primary pocket (top right). Reduced FBXL3 activity can explain enhanced CRY1 mutant repression since the mutants will likely reduce FBXL3 binding resulting in higher CRY1 concentrations in cellulo. Interactions of the mutated positions and interactions are shown for H355 (ordering of FBXL3 C-terminus), Q407 (water bridged FBXL3 backbone contact) and R421 (salt bridge to FBXL3 E368) (bottom). The residue numbering of CRY2 is shown in smaller font. E: Purification result for the CRY1 mutants. All proteins are >95% pure after SEC (fin/SEC lane). CatEx enriched CRY1 further after IMAC.

4.1.4.2 Isolation and Cloning of CRY-m CDS

Expression and purification of mouse, human and monarch butterfly (*Danaus plexippus*) repressors CRYs is established in the lab. Mouse and human CRYs have very high sequence identity, therefore mouse CRY-m was taken as a representative for mammals. The CRY-m:BMAL TAD interaction should be tested for a number of species. The BMAL peptides were purchased but for the CRY-m genes the traditional approach was taken: RNA was extracted from *Apis mellifera* (heads of three workers), *Aedes aegypti*, *Temnothorax nylanderi* and *Tribolium castaneum*. Crude total RNA was purified, and reverse transcribed to cDNA using oligo(dT) primers or gene specific primers for full-length or PHR. Repressor CRY coding sequences (CDS) were amplified from cDNA. *Anopheles gambiae* CRY2 was amplified from pAc5.1/V5-His. Primers were used to yield full-length and PHR constructs. Some insect CRYs have an unfolded N-terminus of unknown function similar to mouse CRY2 (Table 1, Figure 7). N-terminal deletions are denoted ΔPHR. The results of CDS amplification are shown in Figure 16 B.

The CDS of *T. nylanderi* repressor CRY-m could not be amplified, independent of cDNA primer used. The annotation might not have been correct. No full-length (fl) *A. aegypti* repressor CRY was amplified, but PHR and ΔPHR worked. For *A. mellifera* and *A. gambiae* all repressor CRY construct could be amplified. *A. gambia* doesn't have a N-terminal extension so only fl and PHR were cloned. The *A. gambiae* CRY-m C-terminal tail is about the length of the PHR. Only *T. castaneum* repressor CRY PHR could be amplified no matter which primer was used for cDNA synthesis. For some reason the full length reverse primer yielded cDNA but then the full length CDS could not be amplified even though the 5' end apparently was intact since the PHR PCR worked. Full length and PHR CDS share the same 5' end. *A. aegypti* PHR could only be amplified from oligo(dT) cDNA (Figure 16 B). Gene specific primers were personally expected to work better because of higher specificity potentially yielding a more resource efficient cDNA synthesis by avoiding reverse transcribing every isolated polyA RNA. Contrary to expectation amplification success was best with oligo(dT).

CRYs were cloned into pCoofy27 by Gibson assembly, placing it under control of the polyhedrin promoter and adding a N-terminal 3C-cleavable His_7-tag. The expression cassette was transposed into recombinant baculovirus genome in DH10 Bac YFP and virus produced by transfecting baculovirus into insect cells. *A. aegypti* repressor CRY-m isoforms were found. Experiments were conducted with the one closer to the database form (Figure 17). Only the CRY-m PHR was successfully cloned for all species (but *T. nylanderi*), further experiments were performed on PHR for better comparability.

4.1.4.3 Expression and purification

CRY-m PHRs were expressed in 2 L of insect cells from P1 virus. Proteins were purified by affinity chromatography (IMAC) followed by size-exclusion chromatography (Table 26). The mosquito (*A. aegypti, A. gambiae*) CRYs yielded the most protein. The purification of *T. castaneum* resulted repetitively in a loss of protein during the concentration step resulting in too low yields. *A. mellifera* CRY-m performed ok. Some purified samples are shown in Figure 16 D. The sequences are shown in Figure 17.

4.1.4.4 Affinity measurements

Affinities of mouse and insect CRY-m PHRs to mouse and insect BMAL1 peptides was determined by fluorescence polarization (FP). The mouse BMAL1 TAD and *D. plexippus* (monarch butterfly) BMAL TAD were used (mTAD and dpTAD). Shorter peptides were used to identify sequence determinants of binding affinity representing the α-helical (mouse: mAlpha, insect: inAlpha, Monarch: dpAlpha) and C-terminal (mouse: mCterm, insect: inCterm, Lepidoptera: lepCterm, Monarch: dpCterm) binding epitopes. All BMAL peptides have a N-terminal fluorescein (FAM) label. The sequences and result of the measurements is shown in Figure 16 A and C, respectively.

Interestingly, all CRYs but *Danaus plexippus* CRY-m show a 10-fold higher affinity to mouse BMAL1 TAD than to *D. plexippus* BMAL TAD (dark blue background). *Danaus plexippus* CRY-m only shows a 5-fold higher affinity to mouse BMAL1 TAD. The C-terminal binding epitope shows an interesting profile of decreasing affinity with every step from mouse to insect to lepidoptera to monarch (light blue background). The monarch butterfly CRY-m shows no difference in binding affinity between the Lepidoptera and monarch BMAL C-terminal peptide. The main difference between the C-terminal peptides is that with every step from mouse towards lepidoptera one negative charge is lost (D to G and D to N exchange). The last step lepidoptera to monarch comes with the F to I exchange which doesn't negatively impact CRY-m:BMAL TAD affinity in the monarch butterfly protein. The addition of the C-terminal proline alone doesn't change affinity (data not shown). Affinities of the α-helical peptides don't follow a clear trend. Exchanges (S to G and D to G) do not bring about recurrent affinity changes. Affinity differences between mouse and monarch BMAL TAD α-helical motifs are not as strong

as the differences for the C-terminal motif. The differences in affinity between mouse and monarch BMAL TAD can be explained by the clear preference for negatively charged C-terminus.

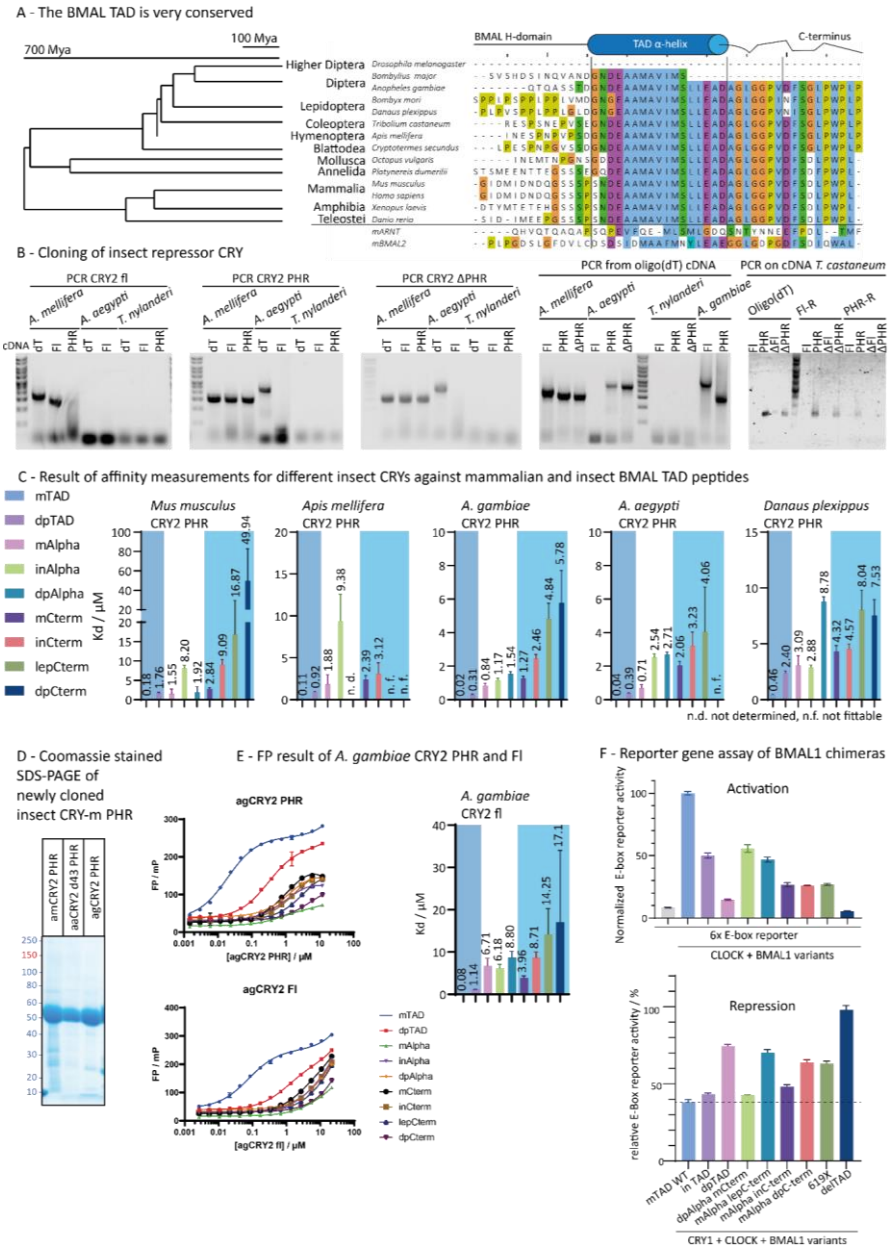

Figure 16: Result of BMAL TAD variant and insect CRY-m interaction. A: Multiple sequence alignment of BMAL across the animal kingdom, mouse BMAL2 and ARNT and phylogenetic tree[163] *showing the*

high conservation of the BMAL TAD sequence. B: Result of RNA extraction and cDNA synthesis with different primers (oligo(dT), full-length reverse and PHR reverse). No CRY-m sequence could be amplified for the ant Temnothorax nylanderi. Full length CDS could only be isolated from the honeybee, Apis mellifera. Oligo(dT) cDNA gave the best results. C: Result of the affinity measurements between insect CRY-m and BMAL TAD peptides. A low K_D value indicates high affinity. The binding profiles are very similar. D: Purification result of the newly cloned insect CRY-m PHRs. Apis mellifera CRY-m is the least pure. E: FP curves for BMAL TAD: Anopheles gambiae CRY-m fl/PHR interaction. The PHR binds with 2-4-fold affinity. F: Result of a reporter gene assay using BMAL TAD chimeras. The monarch BMAL TAD (dpTAD) fails to activate transcription in similar manner as a TAD deletion mutant (delTAD). The insect and monarch C-terminus (inCterm, dpCterm) also reduce activation and are comparable to a C-terminal deletion (619X) which is known to be phase advanced by three hours. The grey bar signifies transfection of only reporter and therefore the background noise. N=3, mean and SEM are shown.

4.1.4.5 Repressor binding preference is very conserved

Species specific binding preference was expected. Surprisingly, the binding profiles look very similar between the different CRY-m proteins whose common ancestor dates back more than 700 million years. All CRY-m proteins bind with higher affinity to C-terminally negatively charged BMAL TADs compared to uncharged C-termini. This preference is the strongest in mouse CRY2 with a 6-fold loss in affinity (mouse to lepCterm). The presence of the phenylalanine appears also to be important for binding as its absence decreases affinity a further 3-fold in the mouse. These differences are less pronounced in the insect CRYs. The affinities for the TAD are submicromolar in all CRYs with the lowest affinity found in monarch CRY-m. Monarch CRY-m shows the smallest window of affinities in our BMAL set due to the relatively low mouse BMAL1 TAD affinity and apparent indifference to the C-terminally F-I exchange. Monarch CRY-m seems to have compensated for the BMAL F-to-I exchange indicating co-evolution of the two proteins. Mouse CRY2 seems to rely the strongest on the C-terminal negative charges.

4.1.4.6 Effect of repressor CRY tail on BMAL TAD binding

The role of the CRY-m tail in BMAL TAD binding is not well understood. The very long *A. gambiae* tail (Figure 27) reduces affinity 2 to 4-fold as shown in Figure 16 E. The *Danaus plexippus* CRY-m tail on the other hand has a less pronounced effect on BMAL TAD binding as shown in Figure 18 B. The tail is also quite long as shown in Figure 27. The human CRY1 tail was shown to reduce PER2 binding but not BMAL1 TAD binding and was proposed to work autoinhibitory[62]. The human CRY1 tail is much shorter than the butterfly/moth and mosquito CRY-m tail as shown in Figure 27. An effect on BMAL TAD:CRY-m affinity might only get relevant at long tails or could alternatively or additionally be composition dependent.

4.1.4.7 Insect TADs lack activation and repression in a cellular model

The effect of BMAL TAD variation was studied in a cellular model by a reporter gene assay, again in collaboration with the Kramer lab. Mouse BMAL1 chimeras with insect TAD, monarch (*Danaus*

plexippus) TAD and α-helix and C-terminal fragments were cloned to reflect as good as possible the peptides tested in the fluorescence polarization assay. Additionally, a C-terminal deletion mutant (619X) and a TAD deletion (delTAD) were tested. The 619X mutant is known to result in 3 h shorter periods[71]. Activation and repression by mouse CRY1 were tested. Results are shown in Figure 16 F. CRY1 plasmid amount was chosen to reach at 50% inhibition to avoid floor effect. Repression was normalized to activation within each BMAL1 chimera.

The insect TAD (inTAD) is only able to activate transactivation by 50% compared to mouse BMAL1 WT (α S->G, C-term D->G, additional P). Activation is almost absent in the monarch TAD (dpTAD) and absent in a TAD deletion mutant (delTAD). Replacement of the mouse TAD α-helical motif with the respective monarch motif (S->G, D->G substitution) also yields about 50% transactivation as in insect TAD with the same substitutions but at different positions. The lepidoptera C-terminus (lepCterm, D->N, D->G) also yields about 50% transactivation. Interestingly, transactivation is further reduced if only the insect C-terminus instead of the whole inTAD is placed in the mouse BMAL1 TAD (mAlpha inC-term, D->G) resulting in about 30% activity. This activity is very comparable to the effect of the monarch C-terminus (mAlpha dpC-term) and the 619X deletion mutant. Insect and monarch BMAL TAD C-terminus don't increase activation compared to the absence of the C-terminus. 619X still contains D and F. Uncharged amino acids in the C-terminus negatively affect transactivation while the α S->G substitution increases transactivation. Further, the α D-> G substitution reduces transactivation as well. The known activator is CBP/p300 binding via KIX and TAZ domains. Other interaction partners of the TAD are possible but unknown.

The used amount of CRY1 reduced BMAL1 WT activity to ~43%. Repression cannot be usefully calculated in the delTAD mutant since the activation was already below control. Slightly less repressive potential as for the WT was observed for the inTAD, dpAlpha and even more in inC-term. Repression is negatively affected in the lep C-term, dpC-term, 619X and dpTAD

The repression data perfectly reproduces the effects observed in the fluorescence polarization assay for mouse CRY1:BMAL TAD interaction.

Co-activator recruitment is more sensitive to changes in TAD than CRY-m. Substitutions in the C-term negatively affect activation while substitutions in the α-helix can increase activation to the point of rescuing C-term (inC-term vs inTAD) effects or negatively affect activation themselves (dpAlpha and probably only D->G).

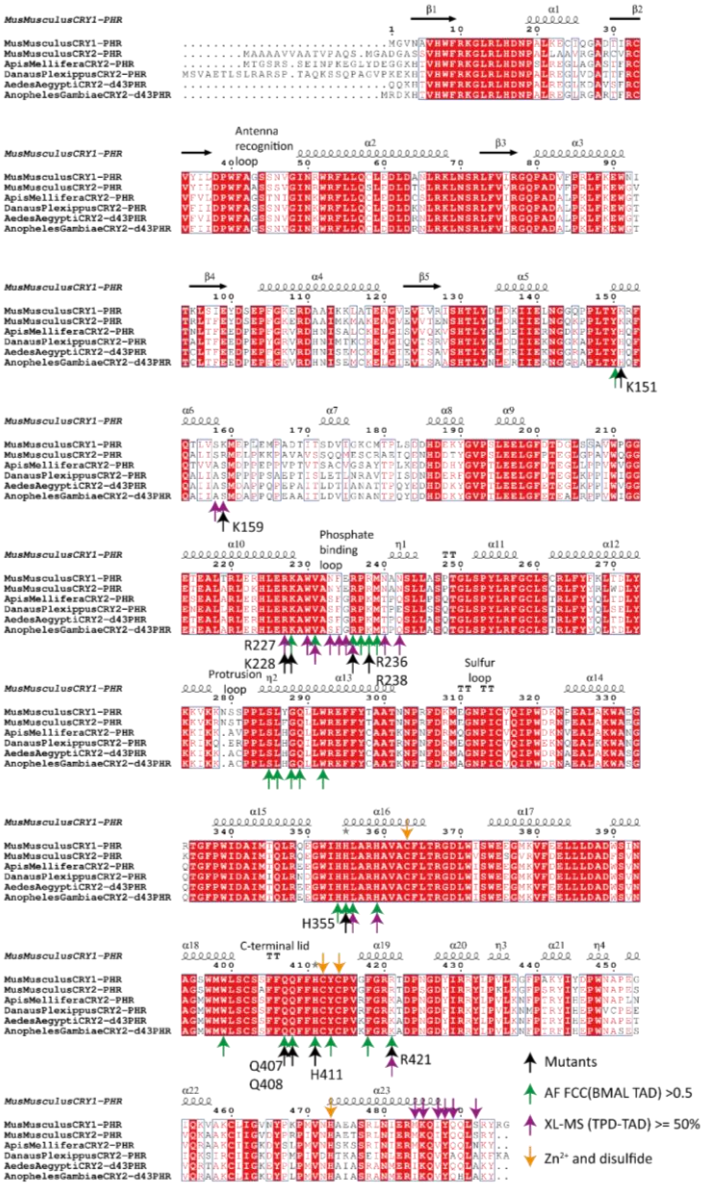

Figure 17: Alignment of the CRY-m used. The CRY-m used herein are shown in a multiple sequence alignment. Mutated positions are indicated by black arrows. Positions predicted by AlphaFold to interact with the BMAL1 TAD with an FCC over 0.5 are shown in green. Positions that occurred in at least half of the TPD-BMAL1 TAD samples are highlighted by magenta arrows. Zn^{2+} interface residues and disulfide bond as well loop annotations are given as reference. XL-MS and AlphaFold cluster at the phosphate binding loop, which is usually not resolved in structures due to its flexibility. The CRY1 R236+R238 mutant that yielded expected result in FP and reporter gene assay is located here.

4.2 Competitive binding of PERIOD and BMAL1 TAD to repressor CRY-m

The repressive branch of the central loop consists of CRY-m and PER, which form a heterodimer in the cytosol and translocate into the nucleus. The CRY:PER complex binds to BMAL and CLOCK over CRY-m via the BMAL TAD as described and shown above. CLOCK is thought to bind to the CRY-m secondary pocket via its PAS-B domain, HI-loop tryptophan[35,92,103]. A regulation of these interactions by PER2 was shown by Fribourgh, Srivastava et al for mouse CLOCK:CRY[96]. PER2 enhanced the affinity of CRY2 to CLOCK to similar levels as CRY1:CLOCK, equalizing their affinity as shown in Table 1. PER2 is proposed to increase flexibility of the CRY2 serine loop, while making the loop in CRY1 more structured .[96] In the same publication Fribourgh, Srivastava et al show CRY1/2 PHR:PER2 complexes have almost identical affinity to the BMAL1 TAD as apo CRY1/2 PHR. This is surprising since the PER2 CBD wraps around the CRY1/2 C-terminal coiled-coil (CC) helix[98,99], identical with the proposed BMAL1 TAD binding site[71,83]. While PER2 was shown to influence CRY:CLOCK affinity the CRY-m:BMAL TAD interaction should be unperturbed even though the binding sites overlap?

4.2.1 The case of monarch butterfly CRY2:PER

Similar to Fribourgh, Srivastava et al [96] the affinity was measured of monarch butterfly apo CRY-m, apo CRY-m PHR and CRY-m PHR:PER complex to a fluorescently labeled monarch BMAL TAD peptide by fluorescence polarization (FP). The expression and purification of monarch butterfly period proteins was optimized in the bachelor thesis of Til Wanner under my supervision. Monarch butterfly PER is soluble in complex with CRY-m in succinic acid, phosphate, glycine (SPG) buffer at pH 7.5 as determined by thermal shift assay (TSA).

CRY-m fl, CRY-m PHR and the CRY-m PHR:PER complex were therefore purified and FP assay carried out using SPG buffer (Table 26, Figure 18).

4.2.1.1 Monarch CRY-m:BMAL TAD affinity measurements

Affinities were measured by fluorescence polarization using fluorescein (FAM) labeled peptides representing the monarch BMAL TAD, the α-helical and C-terminally binding motifs within the TAD according to [71] and the mouse BMAL1 TAD C-terminus (Figure 18 B).

Monarch CRY-m and CRY-m PHR have very comparable affinities indicating only a small negative to no influence of the tail on BMAL TAD binding contrary to the result observed in *A. gambiae* CRY-m. The effect of mouse CRY1 tail on BMAL1 TAD binding is also negligible (data not shown). The affinity of CRY-m to monarch BMAL TAD C-terminus (dpCterm) was not measured.

Kommentiert [CM2]: delete

On the other hand, monarch PER reduces CRY-m:BMAL TAD affinity 4-5-fold for each BMAL TAD peptide tested. This result is in stark contrast to the literature available for mouse PER2 influence on

CRY1/2:BMAL1 TAD binding[96]. AlphaFold multimer[170,171] predictions were used to get an idea what might be the cause.

4.2.1.2 AlphaFold

Predictions were performed with the sequences of monarch CRY-m_{1-520}, $BMAL_{618-649}$ and in a second prediction additionally monarch $PER_{926-1056}$ as fasta input. These sequences represent the fragments used in the affinity measurement. The highest scoring predictions were superimposed and visually inspected with ChimeraX. AlphaFold multimer predicts the monarch BMAL TAD to bind to the CRY-m primary pocket as expected. PER is predicted to bind CRY-m in a very similar way as mouse PER2 binds to CRY1/2, wrapping around the CC-helix[98,99]. PER is predicted to assume a stable tertiary structure starting N-terminally at the CRY-m secondary pocket to the CRY-m primary pocket into which an α-helix is inserted. Sequence homology to mouse PER1/2/3 is present for the CBD but not for the inserted helix and sequences after the CBD as shown in Figure 18 D. Conservation after the CBD, called the tail, is very low, even mouse PER1/2 lack homology after the CBD. *Drosophila* does not have a CBD and therefore no sequence is shown in this sequence snippet.

It seems like a monarch butterfly specific competition was identified between PER CBD and BMAL TAD for CRY-m binding. Even the potential competing epitope could be identified in the sequence. This model is in good agreement with available data, since mouse PER2 lacks a sequence insert where the monarch PER helix is located.

4.2.2 PER reduces CRY-m:BMAL TAD affinity

Reproduction of the literature presented data was the logical next step to validate that mouse PER2 and BMAL1 TAD can independently bind to CRY1/2. The MSA in Figure 18 D suggested that mouse PER1 would behave similar as mouse PER2. We chose the mouse PER2 fragment used in the CRY1:PER2 crystal structure ($mPER2_{1119-1252}$)[98] because of expression yield, established purification protocol, high yield and most importantly because it is similar to the monarch PER fragment used allowing comparisons. For mouse PER1 my bachelor student Jule Urschel had established a very well working fragment ($mPER1_{1116-1291}$) and a number of shorter fragments so that I could select the closest homolog ($mPER1_{1102-1241}$). A domain scheme of PER1/2 based on the uniprot annotation is shown in Figure 19 A. Annotated are domains, motifs as well as parts of PER2 resolved in the available crystal structures[96,98,99] and the constructs used.

4.2.2.1 AlphaFold

In Figure 19 B the CRY1:PER2 crystal structure (PDB 4CT0[98]) and AlphaFold multimer[170,171] predictions for CRY1 PHR:PER2 CBD:BMAL1 TAD, CRY1 PHR:PER1 CBD:BMAL1 TAD and CRY1 PHR:BMAL1 TAD are shown. The prediction of PER2 agrees to the crystal structure and AlphaFold used the available

structures to build the model. The secondary structure assignment is different but the Cα trace is almost identical. AlphaFold fails to incorporate metal ions as of now and so an unoccupied Zn^{2+} interface between CRY1, providing CH motif (C414, H473), and PER2, providing the CxxC motif (C1210, C1213)[98,99] is found .

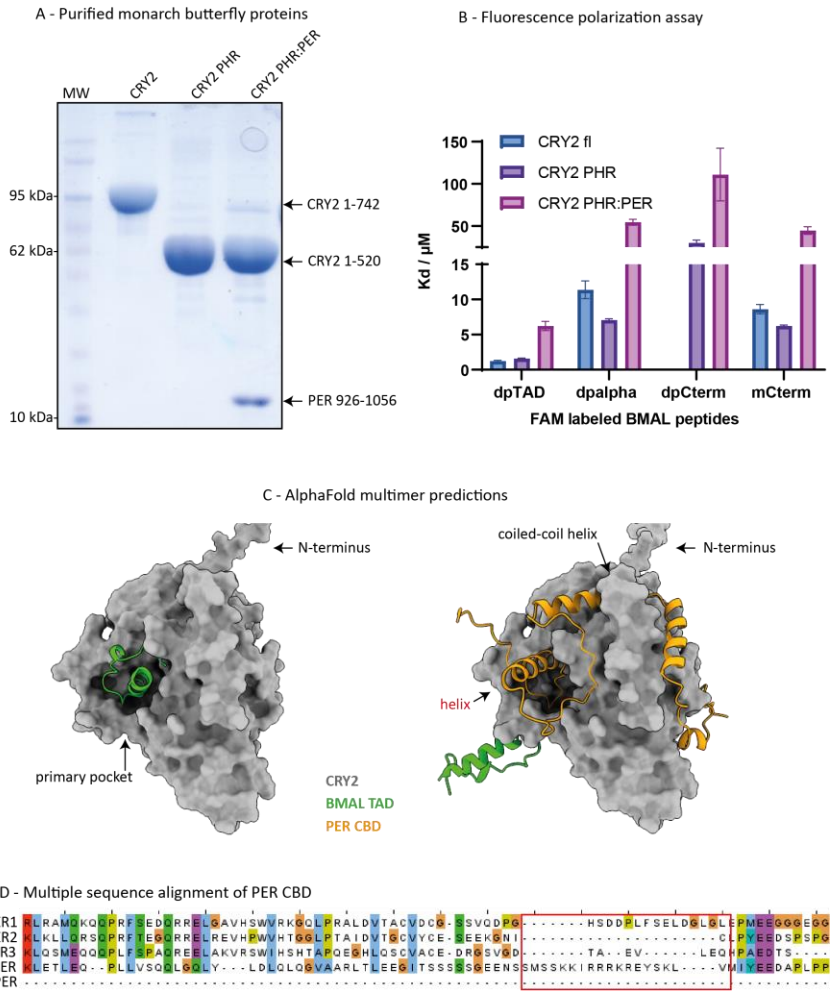

Figure 18: Monarch butterfly PER reduces CRY:BMAL TAD interaction. A: SDS-PAGE of CRY2 and CRY2:PER purification result. B: PER reduces CRY2 PHR:BMAL TAD affinity 4-5 fold in FP. CRY2 tail has only a minor effect. C: AlphaFold multimer predictions suggest that an α-helix after the CBD of monarch PER displaces BMAL TAD in the primary pocket and the responsible sequence is not conserved as shown in a multiple sequence alignment (D).

A zoomed out view of CRY1 PHR:PER1 CBD:BMAL1 TAD and CRY1 PHR:PER2 CBD:BMAL1 TAD is also shown in Figure 19 B on the right. This view allows us to appreciate the length of the unfolded PER1/2 tails in comparison to the CRY1:PER1/2 complex. The PER2 tail is much shorter than the PER1 tail which continues behind the paper plane.

Interestingly, Alphafold predicts independent binding sites for PER2 CBD and BMAL1 TAD, while for PER1 a short unfolded and α-helical part after the PER1 CBD sticks into the primary pocket. The BMAL1 TAD is pushed aside by PER1 compared to the CRY1:BMAL1 TAD prediction. This part is exactly the part found in Figure 18 D in the red box. This suggests that PER1 competes with BMAL1 TAD for CRY1 binding and that the responsible sequence is identified.

4.2.2.2 CRY1:PER1 and CRY1:PER2 complexes have reduced affinity to BMAL1 TAD

The affinity was measured of apo CRY1, CRY1:PER1 and CRY1:PER2 complexes to FAM-labeled BMAL1 TAD by FP as shown in Figure 19 C. A 50-100 fold reduction in binding was observed compared to apo CRY1 by $PER1_{1116\text{-}1291}$ and $PER2_{1119\text{-}1252}$, respectively. This result is in strong disagreement with the literature[96] and AlphaFold prediction (no competition). To validate this, more experiments were needed:

4.2.2.3 PER1 and PER2 break preformed CRY2:BMAL1 TAD complexes

To further characterize this competition PER1/2 were titrated into a preformed CRY2:BMAL1 TAD complex. A competitive FP set-up was used in which CRY2-PHR and FAM labeled BMAL1-TAD are used at constant concentrations of 10 µM and 50 nM, respectively. This should give 60-80% of FP signal compared to saturation, here ~240 mP. PER1/2 are now titrated into this while keeping CRY2 PHR and FAM-BMAL1 TAD concentrations constant. Fluorescence polarization is lost close to equimolar ratio of CRY2 PHR to PER1/2 indicating partially free fluorophore (FAM-BMAL1-TAD). The signal stabilizes around 30 µM PER1/2 and is very close to the signal obtained in the direct measurement shown in Figure 19 C on the left. PER1/2 can bind to CRY2 PHR in a preformed CRY2 PHR:BMAL1-TAD complex, and the resultant CRY2 PHR:PER1/2 complex binds FAM-BMAL1-TAD with reduced affinity. PER1/2 binding is dominant over BMAL1 TAD binding.

PER2 is the stronger competitor than PER1 in both the direct and competitive experiment. The affinity of PER1/2 to CRY1/2 are insignificantly different in the single to double digit nanomolar range or one to two orders of magnitude stronger than the CRY1/2:FAM-BMAL1-TAD interaction a shown in Figure 38. What is the basis for stronger PER2 competition if not a stronger binding to CRY1/2?

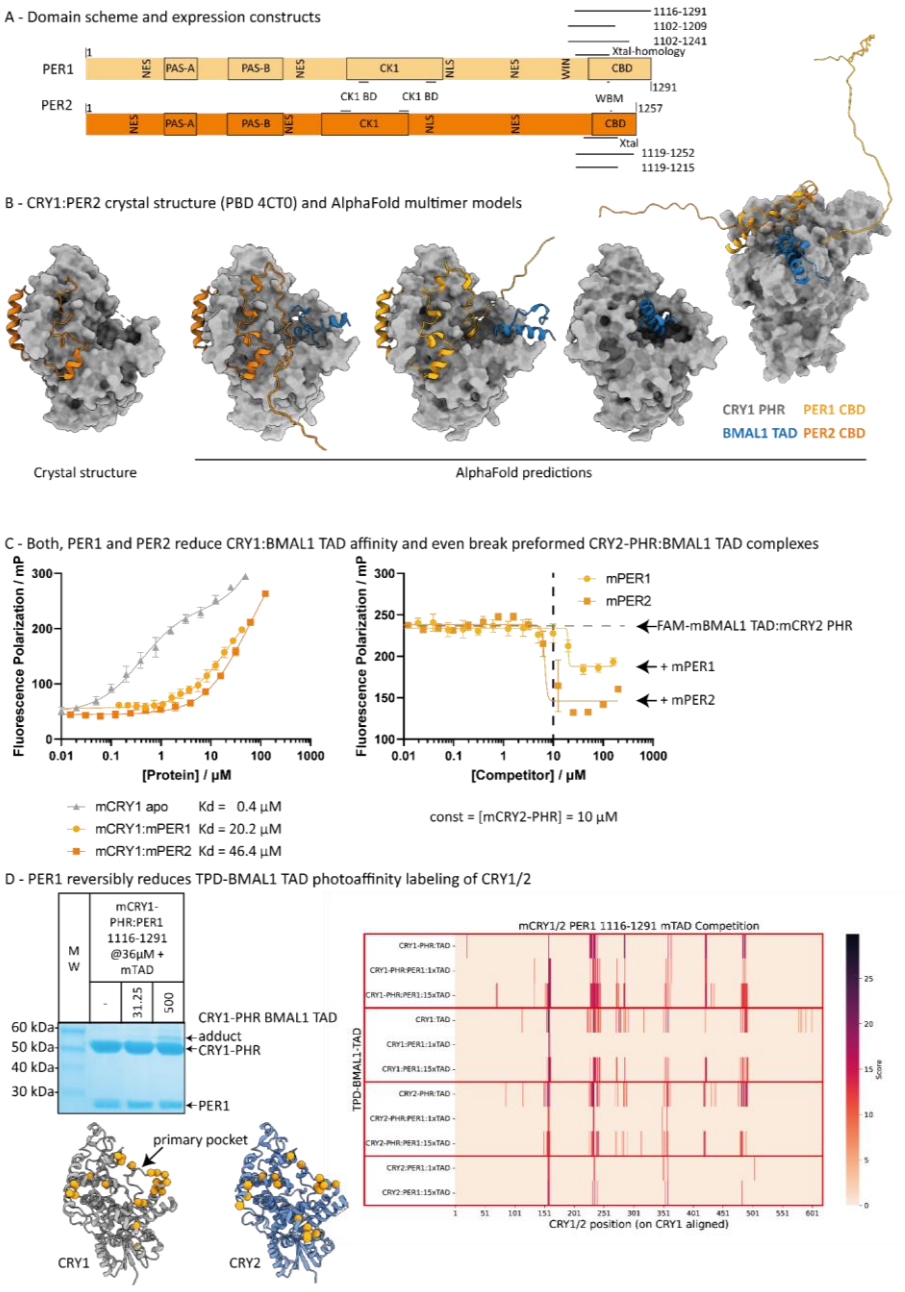

Figure 19: Competition of mouse PER1/2 with BMAL1 TAD for CRY1/2 binding. A: PER1/2 domain schemes based on uniprot annotation. B: AlphaFold multimer predictions. C: PER1/2 CBD reduces CRY-m:BMAL1 TAD affinity in FP. D: PER1 reduces photoaffinity labeling of CRY1/2 by TPD-BMAL1 TAD.

4.2.2.4 PER1 reduces photoaffinity labeling

The established photoaffinity labeling (PAL) was used to test how PAL would also be affected by presence of PER1/2. Since the FAM label can change physicochemical properties, the TPD label which gives us positional data seemed well suited. Additionally, It would tell us if the N-terminus of BMAL - TAD gets into close proximity of PER1/2. Complexes from an earlier ITC experiment (Figure 38). that had been run over SEC and concentrated were used. Complexes at ~35 µM were mixed with ~ equimolar amount and ~ 15-fold molar excess of TPD-BMAL1 TAD and photo-crosslinked.

While an equimolar ratio of TPD-BMAL1 TAD is sufficient to get a visible band shift, no CRY1-PHR band shift was visible in presence of $PER1_{1116-1291}$ (Figure 19 D). At ~15-fold molar excess TPD-BMAL1 TAD is able to overcome PER1 competition and a CRY1-PHR band shift is visible again. Interestingly, the PER1 band never showed a band shift suggesting that the BMAL1 TAD N-terminus and PER1 never get in close contact in a stable conformation and that binding is mutually exclusive or at least flexible at the concentration used. Considering the FP data in Figure 19 C, at ~35 µM more than 50% of a 50 nM FAM probe was bound. In the PAL experiment we have 500 µM TPD-BMAL1 TAD or about 25-fold excess over K_D and should have saturated the CRY1 PHR:PER1 complex. So either BMAL1 TAD displaces PER1 or they can somehow bind together but are not in close proximity.

This experiment was performed with both CRY-m paralogues and their PHRs. The results are shown in Figure 19 D as a heatmap and examples of crosslink positions for apo CRY1/2 plotted as yellow spheres on CRY1 and CRY2 PHR structures. The crosslinks are reduced by PER1, but regained at 15-fold molar excess of TPD-BMAL1 TAD. Additional crosslinks only found in the 15x samples could be unspecific binding at these concentrations. Interestingly, in CRY1 the crosslinks to the CRY1 tail are lost even at high TPD-BMAL1 TAD concentrations. These positions fit to a BMAL1 binding epitope in the CRY1 tail[83]. One can only speculate but perhaps PER1 stays bound to the CRY1 tail. The function of the CRY-m tails is still largely unknown, but we know that CRY1 exon 11, the middle of the tail, binds to the PHR[62] and this interaction seems to be important in delayed sleep phase disorder (DSPD)[57,62,192].

4.2.2.5 Paralogue specific origin of competition

The lengths of the tails are very different in PER1/2 as shown in Figure 19 B on the right. Hypothetically, these parts are dispensable for CRY1/2 association and flexible based on lack of resolution of the PER2 tail in the CRY1:PER2 structure (PDB 4CT0[98]) which included the tail ($PER2_{1216-1252}$). Unfortunately, the minimal CBD constructs $PER1_{1102-1209}$ and $PER2_{1119-1215}$ weren't very soluble by themselves restricting the available methods. CRY1:PER1/2 CBD complexes were purified from co-lysis of insect cell expressed CRY-m and *E.coli* expressed PER1/2. For better comparison the same purification protocol was used for the "CBD+tail" constructs $PER1_{1102-1241}$ and $PER2_{1119-1252}$ and CBD and CBD+tail complexes of one PER paralog were purified in parallel. Affinities of these complexes to FAM-BMAL1-TAD were determined

by fluorescence polarization as shown in Figure 22 A. The data is plotted on a log scale and fold change is also plotted. Affinities are also given in Table 28.

The PER1 CBD reduces affinity 6-fold while the presence of the tail reduces affinity even ~62-fold. Surprisingly, the PER2 CBD alone reaches a similar reduction (~43-fold) while together with the tail a ~190-fold reaction in affinity of CRY1 to BMAL1-TAD is achieved. Even though the reduction in affinity by PER2 is large, it is still evident, that the PER1 tail has a stronger relative effect than the PER1 CBD (K_D(CBD)/K_D(CBD+tail) ~ 10 > K_D(apo)/K_D(CBD) ~6) while for PER2 the reverse is true. The PER2 CBD has a stronger effect on affinity reduction than the tail (K_D(CBD)/K_D(CBD+tail) ~ 4 < K_D(apo)/K_D(CBD) ~43). In a paralogue comparison these differences become even more apparent: the PER1 tail further reduces affinity 7-fold while the PER2 tail is about half as potent with 4-fold. The PER1 CBD only reduces affinity 6-fold while the PER2 CBD reduces 43-fold.

What effect will these differences have in full-length proteins in the context of repression of BMAL1:CLOCK mediated transcriptional activation?

Table 28: PER CBD and CBD+tail constructs used. PER1/2 construct borders are given and the resultant K_D of respective CRY1:PER1/2 complexes to FAM-BMAL1 TAD is shown. All complexes were purified by co-lysis of cell pellets from individual expressions.

FP	**Titrand**		**Kd (SD)/µM**
Probe	mCRY1 fl….	construct	
FAM-BMAL1-TAD	Apo	------	0.338 (0.041)
	PER1 CBD	$PER1_{1102-1209}$	2.080 (0.663)
	PER1 CBD+tail	$PER1_{1102-1241}$	20.980 (0.690)
	PER2 CBD	$PER2_{1119-1215}$	14.593 (1.810)
	PER2 CBD+tail	$PER2_{1119-1252}$	64.220 (12.686)

4.2.2.6 PER1 increases BMAL1:CLOCK transcription activation

Reporter gene assays were performed in a collaborative effort in the lab of Prof. Achim Kramer (Charité Berlin) by Astrid Grudziecki. A tandem E-box promoter driving expression of firefly luciferase was used. Expression of BMAL1 and CLOCK lead to enhanced luminescence (expression), while expression of CRY1 reduces BMAL1:CLOCK mediated luciferase activity (transcription repression). Expression of CRY1 and PER1/2 should elevate BMAL1:CLOCK mediated reporter activity compared to CRY1 alone because the CRY1:BMAL1 TAD interaction would be reduced, based on the results shown above.

Floor and ceiling effects have to be avoided and therefore the assay has to performed with plasmid amounts/protein expression levels that are limiting signal-to-noise ratio while working in the repressed state of the system (CRY1) further reducing signal-to-noise. Results of the reporter gene assay are

shown in Figure 22 B,C as % reporter activity, normalized to maximal BMAL1:CLOCK activity. All data from Figure 22 B,C is the result of the same assay, but plotted differently for better understanding.

As expected BMAL1:CLOCK drive expression/activity of firefly luciferase, while CRY1+PER1 shows elevated reporter activity compared to CRY1. Interestingly, CRY1+PER2 reduced reporter activity further than CRY1 alone (Figure 22 B).

Next the domains used in the biochemical assays were exchanged and deleted to see if the effects are also present in the cell or artifacts of working with isolated proteins (Figure 22 C). As expected PER1 Δtail even slightly reduces BMAL1:CLOCK activity, while a chimera of PER1 core with the competing PER2 CBD increases BMAL1:CLOCK activity relative to PER1 Δtail. Deleting the PER2 tail (PER2 Δtail) on the other hand only has a small effect, while exchanging the PER2 CBD with the PER1 CBD does not change activity compared to PER2 Δtail in the background of PER2 core. In the PER2 core and Δtail context the PER1 and PER2 CBD seem indistinguishable.

The competitive effect of the PER2 CBD can be appreciated when PER2 CBD is placed in PER1 (PER1 core PER2 CBD PER1 tail) by comparing it to PER1 WT, here PER2 CBD reduces CRY1-mediated repression as expected. On the other hand, exchanging the PER1 tail for the PER2 tail (PER1 core PER2 CBD PER2 tail) enhances repression compared to PER1 WT, rendering it to behave neutral (repression neither reduced or enhanced). This shows again how strongly the PER1 tail reduces repression.

In the context of PER2 core, the PER1 tail (PER2 core pER1 CBD PER1 tail) reduces repression, while the PER1 CBD (PER2 core PER1 CBD PER2 tail) behaves indistinguishable from PER2 CBD (PER2 WT).

4.2.2.7 Core supersedes CBD and tail effects

To sum up the PER1 tail reduces repression, while the effects of PER2 CBD are less profound and best seen in the PER1 core context. The PER2 core enhanced repression in every chimera. The PER1 core on the other hand never enhanced repression. The main difference between PER1 and PER2 therefore lies in their core domains (~85% of the protein). As shown in Figure 19 A the domain organization is very similar between PER1 and PER2. The sequence homology is much lower outside the folded domains, PAS-A, PAS-B and CBD. Important binding sites like the CK1 binding sites and FASP region are again conserved. Many factors on the transcriptional, translational, and post translational level could already influence protein abundance. Recruitment of CK1 or other factors might be different even in conserved binding sites due to minor changes. Nevertheless, evidence was provided that PER1/2 do NOT work as co-repressors for the CRY1 mediated repression of the BMAL1 TAD but instead reduce such an interaction.

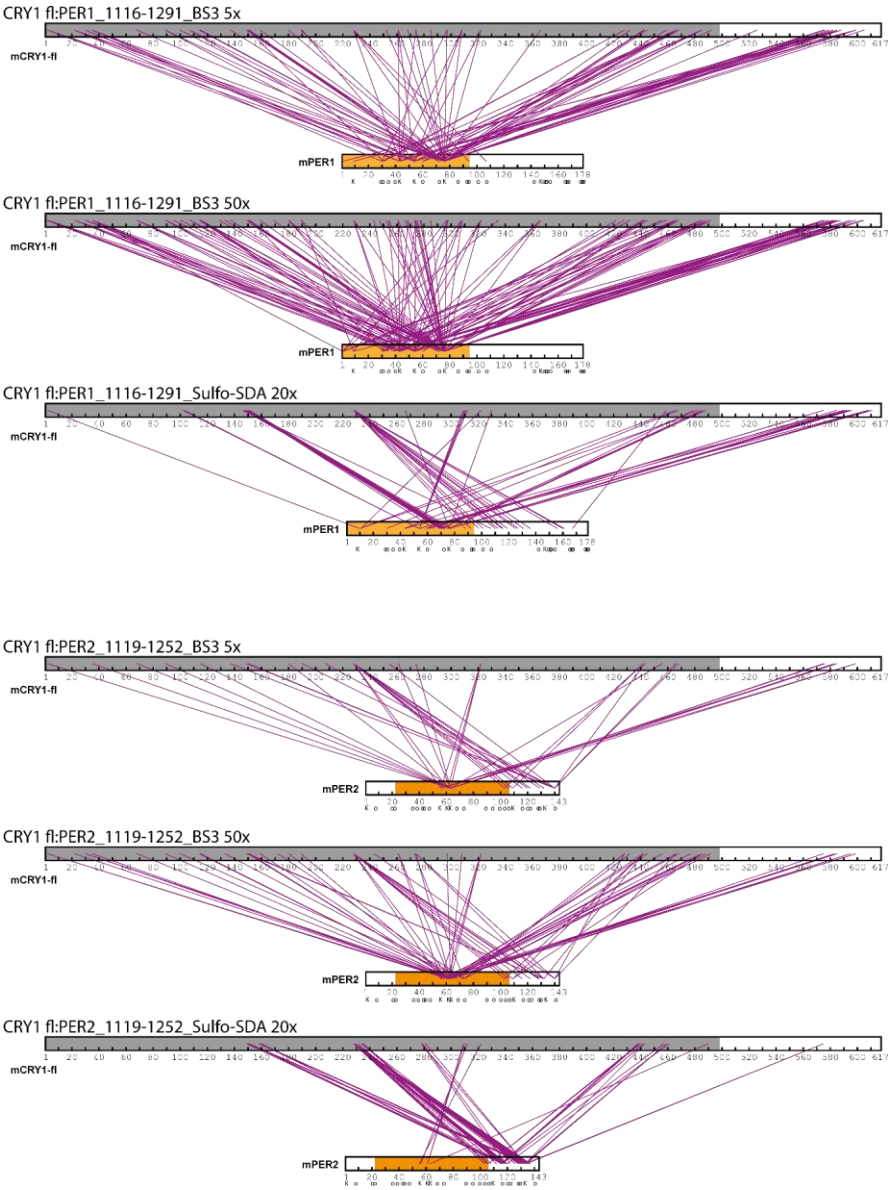

Figure 20: CRY1:PER1/2 crosslinks. Top: PER1$_{1116-1291}$ crosslinks, bottom: PER2$_{1119-1252}$ crosslinks. First from top: crosslinks of 5x molar excess of BS3, middle crosslinks of 50x molar excess of BS3, bottom crosslinks of 20x molar excess of sulfo-SDA. CRY-m PHR in grey, PER CBD in orange. PER residues able to react with NHS are indicated, K: lysine; o: serine, threonine, tyrosine; the N-terminus is not labeled. PER1 tail does not react with BS3, The PER2 tail appears to be more ordered because it crosslinks with BS3 to CRY1 contrary to the PER1 tail. PER tails crosslink with the phosphate loop (CRY1$_{230-240}$). The same CRY1 region that crosslinked to TPD-BMAL1-TAD, CRY1 tail exon 12, crosslinks to PER CBD α4.

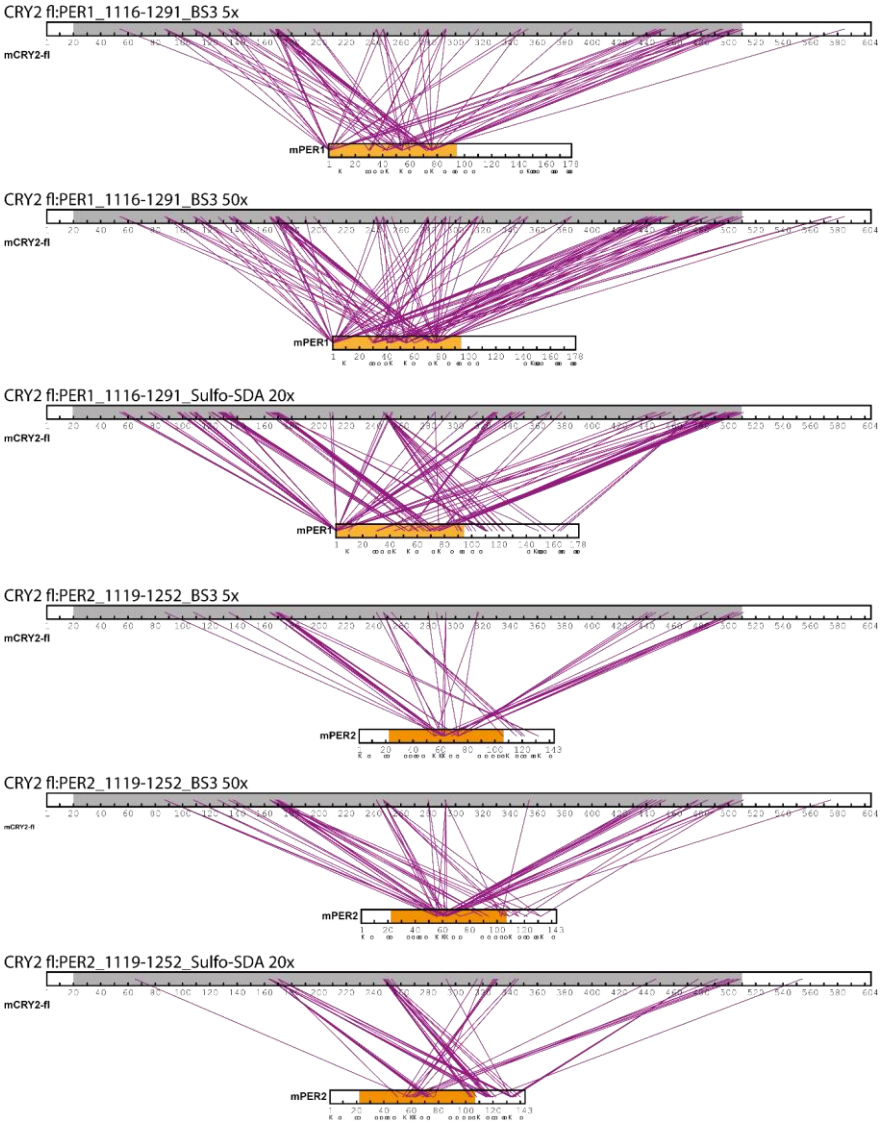

Figure 21: CRY2:PER1/2 crosslinks. Top: PER1$_{1116-1291}$ crosslinks, bottom: PER2$_{1119-1252}$ crosslinks. First from top: crosslinks of 5x molar excess of BS3, middle crosslinks of 50x molar excess of BS3, bottom crosslinks of 20x molar excess of sulfo-SDA. CRY-m PHR in grey, PER CBD in orange. PER residues able to react with NHS are indicated, K: lysine; o: serine, threonine, tyrosine; the N-terminus is not labeled. PER1 tail does not react with BS3, The PER2 tail appears to be more ordered because it crosslinks with BS3 to CRY2 contrary to the PER1 tail. PER tails crosslink with the phosphate loop (CRY2$_{250-250}$). The CRY2 tail exon 11 crosslinks to PER CBD α4 but much less often than the CRY1 tail.

4.2.2.8 The CRY1 tail binds to PER α4

CRY-m paralogs and PERs have disordered and divergent tails. Crosslinking was used to identify their potential binding sites even if these interactions are not as stable as necessary for other structural biology methods. Two different crosslinkers were used: The homobifunctional crosslinker BS3 with a long linker between the two NHS groups and heterobifunctional sulfo-SDA with a short linker between NHS and diazirine. NHS crosslinking is slower and considered less specific while the UV-indued carbene diradical of the diazirine reacts very fast and allows to even identify transient interactions. Complexes of CRY1 and CRY2 full-length with $PER_{1116-1291}$ and $PER_{1119-1252}$ were crosslinked with BS3 at 5- and 50-fold molar excess and sulfo-SDA at 20-fold molar excess. Results are shown in Figure 20 and Figure 21.

The BS3 data with PER1 shows many crosslinks, probably because it has more primary amines, the main target site, than PER2 (5 instead of 3). The comparison with sulfo-SDA shows that the PER1 tail only crosslinks in presence of the faster reacting sulfo-SDA but not BS3. The PER2 tail on the other hand crosslinks in presence of either crosslinker. The PER tail crosslinks originate mostly from the CRY-m phosphate binding loop ($CRY1_{230-240}$, $CRY2_{250-260}$, Figure 17). The CRY1 tail crosslinks mostly to the PER α-4 helix with PER1 and sulfo-SDA or PER2 with either crosslinker. The CRY1 tail crosslinks all come from the exon12, where also TPD-BMAL1 TAD crosslinked to and from which the BMAL1 TAD binding epitope P2 was generated[83].

4.2.2.9 Differences of PER1 and PER2

Zooming back into the differences in PER1 and PER2 competition with BMAL1 TAD for CRY1 binding. There must be important differences between PER1/2 CBD and between tails. What are the differences between PER1 and PER2 tails? Obviously, the length and composition (Figure 22 D). The PER1 tail is longer than the PER2 tail but even with the shorter PER1 C-terminus used in the experiments in Figure 22 A ($PER1_{1102-1241}$) the effect is very much measurable. The PER1 tail is very rich in glycine, rich in small hydrophobic residues and has multiple negatively charged patches. PER2 has a rather unspectacular looking tail. The PER1 tail appears to be more flexible than the PER2 tail as shown by crosslinking (Figure 20, Figure 21). The BMAL1 TAD is also negatively charged and the interaction with the primary pocket seems to be partially dependent on electrostatic interaction and steering. Therefore, the PER1 tail might partially occupy the primary pocket and or at least block electrostatic steering due to its flexibility and negative charges.

What is the difference between PER1 and PER2 CBD responsible for the PER2 CBD competition? The sequences are almost identical (Figure 22 D), the interaction with CRY1/2 has insignificantly different K_D values (Figure 38) and based on the sequences, available crystal structures of PER2[96,98,99] and homology models one can build from it, it is rather mysterious what leads to the observed effects. So

far, structural evidence is lacking that might reveal unexpected details. Unfortunately, by the time the importance of such a structure was realized the necessary time was unavailable.

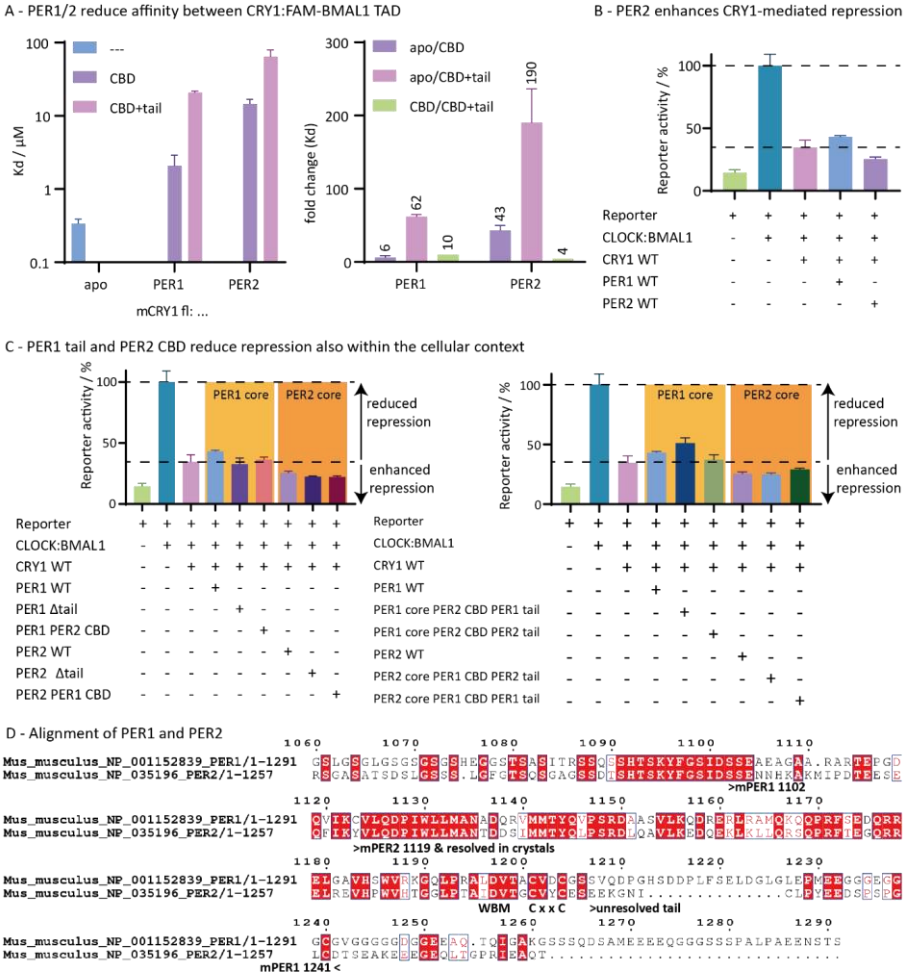

Figure 22: Basis of competition by PER1/2. A: FP based affinity measurements show PER1 tail and PER2 CBD repress CRY1:BMAL1 TAD association(n=2-3). B: PER1 and PER2 reduce and enhance CRY1-mediated repression of CLOCK-BMAL1 activity in reporter gene assays. C: The FP results are confirmed in reporter gene assays, but the PER2 core supersedes effects of the CBD and tail. D: An alignment of PER1/2 hints towards composition as the important difference between PER1 and PER2 responsible for competition.

4.3 Evolutionary analysis of interactions sites in the circadian clock central loop

The elements of the central feedback loop of the circadian clock are very conserved. A statement to be found in every second chronobiology paper because it is true. Even tough multiple circadian clock networks were identified in animals and more suspected[43,50] they usually comprise homologous genes (and proteins) for *cryptochrome* (CRY), *period* (PERIOD), *bmal* (BMAL) and *clock* (CLOCK). Important accessory proteins are usually also present such as E3-ligases[41,42], protein kinases[58,59,114,140,193] and chromatin modifiers[70,76,194,195]. Many studies focus on the presence of the aforementioned genes to study the evolutionary history of the clock or clock components[43,115,169]. Yet some components of the clock come in multiple flavors namely BMAL/CYCLE and PER, while others can be easily misidentified as clock proteins namely ARNT or wrongly assigned due to high sequence identity (ARNT, BMAL, CLOCK). In *Drosophila* BMAL is replaced by CYCLE. The proteins share high sequence identity but are functionally distinct. BMAL has a transactivation domain (TAD) which was studied in detail in the preceding subchapters. CYCLE on the other hand stops right after the PAS-B domain, therefore missing the BMAL TAD as an important site for co-activator recruitment and repression[71]. BMAL/CYCLE homologs missing the TAD will be called CYCLE, while homologs with a TAD will be called BMAL from here onwards, independent on their database name. Further, *Drosophila* PER lacks a CRY-m binding domain. Most parts of the protein are very conserved even compared to mammalian PER, but this important interaction site is missing. These differences are well known in the (biochemical) chronological community but as to my knowledge were not studied systematically in a broader range of animals. *Drosophila* is unsurprisingly the example for those differences here because the *Drosophila* clock works distinctly different from other animal clocks. The questions I wanted to answer are:

- Can we identify species with unknown clock networks?
- Can we identify organisms, clades or lifestyles with specially adapted clocks?
- Do gene presence and interaction site (as a proxy for function) correlate well?

4.3.1 Study design – species set

But the very first question was which species to choose for such an analysis. They should origin from several climate zones, have diverse social lifestyles, lifespans, timing of activity over the day and seasons while their genomes should be annotated well. Insects were chosen because knowledge and material would be available over the GenEvo research training program, and insects fulfill all the above criteria. Which species specifically to choose was a much harder choice and originally a top-down approach was tried. This means data was collected for as many insect species as possible. Valentine Patterson, a bioinformatician and fellow GenEvo PhD student, helped with the bioinformatics part (data collecting, sorting, sequencing expertise). Genom selection criteria were developed and a database for insect genomes, InsectBase2.0[196] was found. Unfortunately, automated annotation of BMAL and CYCLE

didn't work as well as expected and it was realized that incorrect sequences are part of InsectBase2.0. Starting the same approach from a bigger database like the NCBI would have been more complicated and applying the quality criteria would have been much harder. Finally, a species set from a recent study by Kotwica-Rolinska, Chodáková et al (2022)[169] was chosen. The authors were very kind to share their sequences. The gene dataset was extended by BMAL and CLOCK. Collecting the sequences is the most time-consuming step, because much of the data is based on transcriptome shotgun sequencing (TSA) and many sequences used had to be downloaded, translated in the right frame and the protein sequence checked for correctness. ARNT, BMAL1, BMAL2, CLOCK and PER were sometimes found with each other due to PAS domain homology but could successfully be annotated based on percent identity.

Interaction sites were defined based on literature[35,46,71,76,78,79,92,96,98,99,101,121] and many of them are continuous epitopes, allowing analysis by simple multiple sequence alignments including the motif in question.

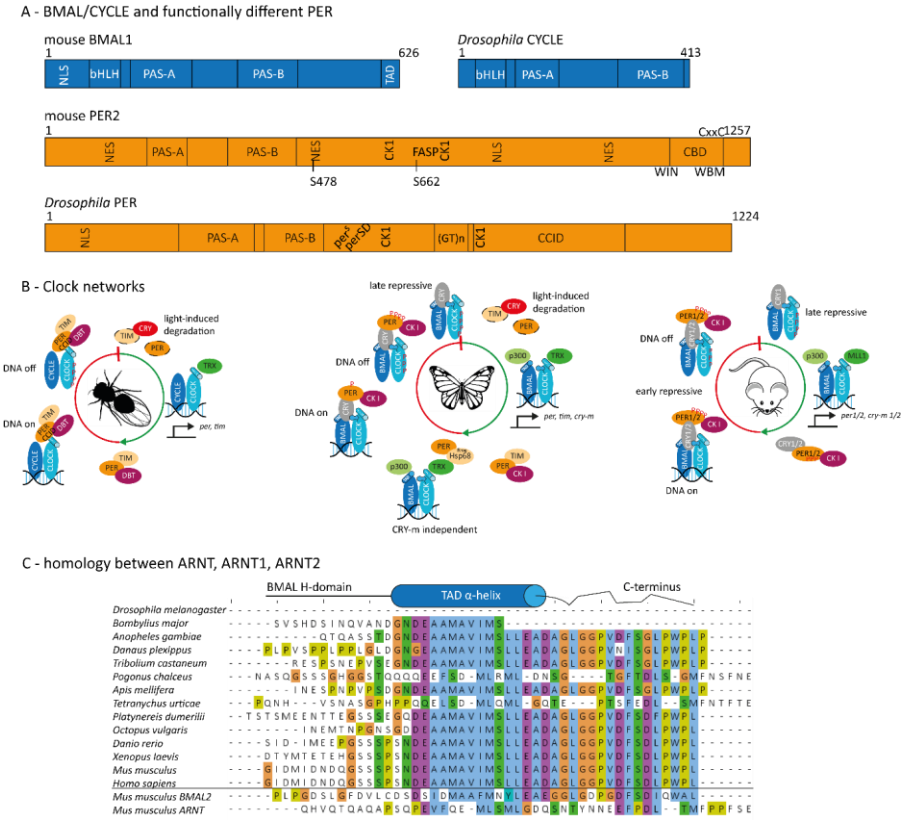

Figure 23: Protein homologs and clock networks. A: Domain schemes of mammalian and Drosophila BMAL/CYCLE and PER show functional differences between the homologs that can be identified based

on the interaction sites. B: The canonical clock network structures from Yuan et al[43] modified to include the most recent literature[78,84,114]. Drosophila on the left, the ancestral clock present in the monarch butterfly and the mammalian clock. The honeybee clock is very similar to the mammalian clock but lacks the gene duplication in cry and per. C: The BMAL TAD is very conserved from insects over octopus to mammals. ARNT and ARNTL/BMAL homologs and paralogs (bottom) can be distinguished by their C-termini.

4.3.2 Interaction site definitions used

The interactions were described in the introduction at length and will be shortly summarized here.

4.3.2.1 BMAL and CLOCK

The interaction between BMAL CLOCK by their bHLH, PAS-A and PAS-B domains is so well conserved that we didn't study it in detail, but we found all three domains in both proteins if the protein was found in the database. The interaction between CLOCK exon 19 and CIPC are well understood on structural level where two exon19 units and one CIPC form a coiled-coil interaction suggesting a necessity for two spatially close CLOCKs for interaction.[79] While CIPC was originally suggested to be mammal specific repressor and not present in the insect lineage[80] some evidence was provided for a homologous protein in insects.[79] CLOCK exon19 and its homologs are also suggested binding sites for the histone methyltransferase MLL1[76] or its insect homolog TRITHORAX[78]. How exactly those interactions look is still unknown. It was suggested that in species lacking a BMAL TAD a polyQ region at the CLOCK C-terminus can recruit chromatin modifiers [50,87].

The BMAL TAD is a well-established interaction site for histone acetyltransferase CBP/p300 recruitment and CRY-m repression by binding site competition.[71,72,83,197].

4.3.2.2 Cryptochromes

Cryptochromes important for this thesis come in two flavors, transcriptional repressors, and photoreceptors. Transcription repressor CRYs were discovered in mammals, while the photoreceptor CRY was discovered in *Drosophila* so until today common names of the two classes are mCRY and dCRY or CRY-m and CRY-d or CRY2 and CRY1. Since I worked during my project with mCRY1/2, agCRY2, aaCRY2, amCRY2, tcCRY2, tnCRY2 and dpCRY2, therefore with different species and numbering due to paralogs, I choose to use CRY-m and CRY-d from hereon.

CRY-m and CRY-d can be distinguished by percent identity and comparison to known CRYs. First CRY-d developed from photolyases and later CRY-m branched off early in the animal lineage.[43,115] The CRY-m were shown to have an internal disulfide bond probably implicated in redox sensing[101] which is very close to the later found Zn^{2+} interface of the CRY-m/PER complex in which CRY-m provides a cysteine and a histidine residue and PER a classical CxxC Zn^{2+} binding site. The disulfide bond formation and Zn^{2+} binding seem to be mutually exclusive, connecting redox metabolism, Zn^{2+} concentration and circadian

functions. [98] The CRY-m of *Anopheles gambiae* has a long Q-tail. Based on the discussion for CLOCK polyQ presence of polyQ stretches in CRY-m tails was investigated.

4.3.2.3 Period

PER also has a PAS-A and a PAS-B domain used for homo- and heterodimerization with PER paralogs which were present in all PERs and not studied in detail here.

Two motifs for interaction with casein kinase 1 (CK1) are described, as well as a phosphorylation site residing in-between with the CK1 consensus phosphorylation motif (S/T)xx(S/T)[198] which is implicated in family advanced sleep phase syndrome (FASP).[59,131,139] Insects have something similar to a FASP region in-between the CK1 binding domains which is less conserved than the chordate FASP but also harbors (S/T)xx(S/T) repeats.

PER can interact with CRY-m if a CRY-binding domain (CBD) is present, located C-terminally of PAS and CK1-BD and followed by an unfolded C-terminal tail. The CBD is intrinsically disordered but acquires an ordered state when binding to CRY-m where it wraps around the PHR and especially CC-helix. PER:CRY-m binding results in higher CLOCK:CRY-m affinity in human CRY2, potentially by increasing flexibility of a loop at the CRY-m secondary pocket.[96,98] The Zn interface between CRY-m and PER was explained above. Recently, the interaction of mouse PER1 and PER2 with a MLL1 complex core component WDR5 was described[199]. PER1/2 interact over a WDR5 binding motif (WBM) and a WDR5 interaction motif (WIN) with WDR5. The WIN motif sits right before the CBD and WBM at the C-terminally end inside the CBD. A complex of CRY:PER1:WDR5 is possible, if the PER1:WDR5 complex is preformed suggesting that the CBD internal WBM is used for WDR5 instead of CRY-m interaction in this complex[199].

Even though the *Drosophila* system is one of the longest and most studied with many available tools, the interaction of *Drosophila* PER with CLOCK and TIM are still not entirely understood. A CYCLE:CLOCK interacting domain (CCID) was identified 20 years ago and surprisingly not much more about the interaction is known[121]. The proposed CCID region is over 200 amino acids long and together with the lack of more detail was excluded from detailed analysis.

Table 29: Investigated interaction sites and the reference sequences used to identify them.

Role	Gene	Site	Interacts with	sequence
activation	CLOCK	Exon19[79]	CIPC and potentially MLL1	QFSAQLGAMQHLKDQLEQRTRMIEANIHRQQE ELRKIQEQLQMVHGQGLQ
		polyQ	Potentially co-activators	Qn, n >= 10
	BMAL	TAD[71]	CBP/p300, CRY-m	SNDEAAMAVIMSLLEADAGLGGPVDFSDLPWP L
repression	CRY-m	Disulfide[101]	Disulfide	HAVA**C**FLTR; QFFH**C**YCPV
		Zn interface[98]	PER Zn^{2+} interface, Zn^{2+}	FHCY**C**PVGF; PMVN**H**AEAS
		Q-rich	Unknown	
	PER	CK1 binding domain A[139]	CK1/DBT	CSYQQISCLDSVIRYLESCNEAAT
		CK1 binding domain B[139]	CK1/DBT	GLTKEVLAAHTQKEEQSFLQKFKEIRKL
		CRY binding domain[98]	CRY-m	IPDTEESEQFIKYVLQDPIWLLMANTDDSIMMT YQLPSRDLQAVLKEDQEKLKLLQRSQPRFTEGQ RRELREVHPWVHTGGLPTAIDVTGCVYCES
		Zn interface[98]	CRY-m Zn interface, Zn^{2+}	CxxC at CBD C-terminus

4.3.3 Circadian network structure is partially order specific

Three dominant network structures are described in the literature and can be found in insects[43].

4.3.3.1 Photoreceptor deficient – Hymenoptera

This system is very similar to the mammalian system. A transcription repressor (CRY-m) and PER together repress CLOCK:BMAL activity. To my knowledge there is no publication describing transcription repression as a two or multistep process in insects.

This network structure is found in Blattodea (cockroaches and termites, 7/10), Phthiraptera (animal lice, n=2), Heteroptera (suborder of true bugs, 4/6 all on land), Hymenoptera (bees, ants, wasps), Polyphaga (beetles, 2/3) (Figure 24).

Chordata use the same structure, but many have gene duplications for CRY-m and PER.

4.3.3.2 Transcription repressor deficient – brachycera

One of the most studied systems because the model organism *Drosophila* is a member of the brachycera subgenus of Diptera (flies). 3/6 brachycera in our species set lack a transcription repressor CRY (CRY-m, Figure 24). Instead TIM-stabilized PER represses CLOCK:CYCLE.

4.3.3.3 Ancestral – monarch

The ancient network structure is the most prevalent in our species set. It is found in Ephemeroptera, Ondonata, Orthoptera, Phasmatodea (all n=1), Blattodea (cockroaches and termites, n =3/10), Thysanoptera (thrips, n=1), Hemiptera (bugs): Sternorrhyncha (aphids, whiteflies, n = 4), Auchenorrhyncha (cicadas, leafhoppers, n=1), Heteroptera (suborder of true bugs, 2/6 all on water). Neuropterida (waxwing lacwings, dobsonfly, n=2/3), Adephaga (suborder of beetles, n=2), Polyphaga (suborder of beetles, emerald ash borer, n=1/3), Trichoptera (n=1), Lepidoptera (butterflies and moths, n= 3), Nematocera (suborder of Diptera, mosquitos, n=2) (Figure 24).

This network structure is also used by Annelida (*Platynereis dumerilii*, marine worm), Mollusca (octopus, sea slugs, oysters), Chelicerata (horseshoe crabs, spiders, mites, ticks), Crustacea and Collembola.

4.3.3.4 Unknown network

In Hemichordata (marine acorn worms) and Echinodermata (sea stars, sea urchins) PER is missing (Figure 24). Reportedly they do have circadian rhythms [200]. Since PER has a very central place in the central loop, one can only speculate how they sustain circadian clock rhythms.

4.3.4 CRY types and interaction sites define network structure

Yuan *et al* (2007)[43] defined networks by presence of the CRY types. I wondered if the CRY type presence/absence would be correlated with their respective interaction sites. For example, in *Drosophila* CYCLE the TAD is absent. Would the TAD be absent in all species lacking a CRY-m? Similarly, would a PER CRY binding domain (CBD) be absent as well?

This would simplify checking the network structure by just looking at PER. BMAL is of course also possible but due to the higher similarity to ARNT this could be misleading.

On the other hand, PER in the brachycera and to a lower extent the ancient system could be optimized for CLOCK:CYCLE interaction. An interaction site could be identified by conservation in species with a brachycera-like network while lacking conservation or presence in species with the Hymenoptera-like network.

4.3.4.1 PER CBD is lost in absence of CRY-m

As expected, PER CRY binding domain (CBD) is lost when CRY-m is lost. Additionally, many cases were identified where PER CBD wasn't found in TSAs even tough CRY-m is present (Table 30). The most significant are the first three, a marine worm, a sea slug and a sea urchin who lost PER entirely according to our database searches and of others[169]. In the other cases the PAS domain core was found and usually the first and often also the second described CK1 binding domain. These PERs therefore look like a *Drosophila* PER with all the major structural domains and interaction sites besides the CBD.

Interestingly, *Panchlora nivea* PER stops right at the beginning of the CBD. Often, only small fragments are identified indicating that something along RNA extraction to TSA analysis went wrong. Generally, absence of evidence is not evidence of absence.

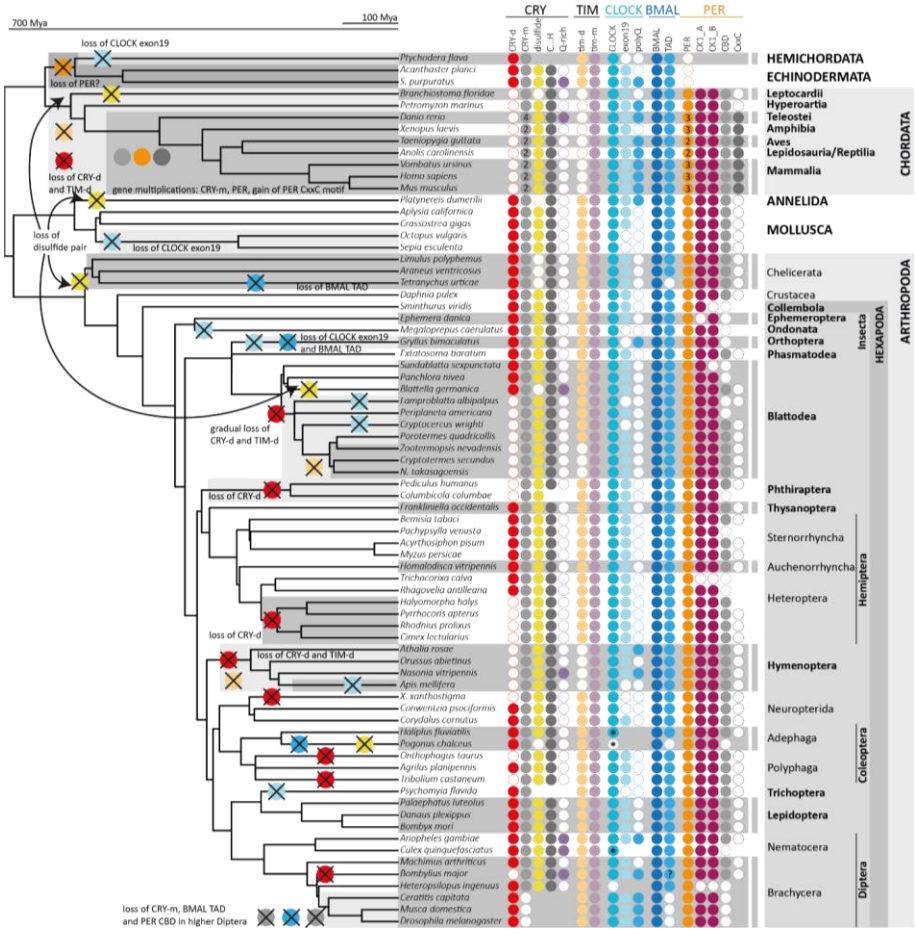

Figure 24: Interaction sites and clock network structures are conserved. Result of sequence search and interaction site analysis were plotted on a corresponding phylogenetic tree (timetree.org)[163] using iTOL[164] and annotated. CRY-d presence was checked (red), while multiple CRY-m (grey) sites were defined (yellow, dark grey, purple). Activators and their interactions sites are shown in shades of blue. PER in orange, PER CK1 interaction sites in purple-red, CBD in CRY-m grey and CxxC in dark grey.

Table 30: Species with a CRY-m but no PER CBD detected.

Phylum	Class	Order	Suborder	Species	TaxID	comment
Hemichordata	Enteropneusta			Ptychodera flava	63121	PER lost
Echinodermata	Asteroidea	Valvatida		Acanthaster planci	133434	PER lost
	Echinoidea	Camarodonta		Strongylocentrotus purpuratus	7668	PER lost
Arthropoda	Collembola	Symphypleona		Sminthurus viridis	109609	Until CK1 BD1 present
	Insecta	Ephemeroptera	Furcatergalia	Emphemera danica	1049336	Many small fragments
		Ondonata	Zygoptera	Megaloprepus caerulatus	263994	Two fragments
		Orthoptera	Orthoptera	Gryllus bimaculatus	6999	CBD missing
		Blattodea		Sundablatt sexpunctata	1603030	Four small fragments
				Panchlora nivea	36955	Stops right at CBD start
				Lamproblatta albiplapus	1080993	Until CK1 BD2 present
				Periplaneta wrighti	6978	Until CK1 BD2 present
		Phthiraptera	Ischnocera	Columbicola columbae	128991	Until CK1 BD2 present
		Hemiptera	Sternorrthyncha	Pachypsylla venusta	38123	Until CK1 BD2 present
			Heteroptera	Trichocorixa calva	1585305	Only PAS domains
			Heteroptera	Rhagovelia antilleana	1921105	Until CK1 BD1 present
		Coleoptera	Adephaga	Pogonus chalceus	235516	Until CK1 BD2 present
		Diptera	Nematocera	Culex quinquefasciatus	7176	Until CK1 BD2 present

4.3.4.2 BMAL TAD is lost in absence of CRY-m

In all three species lacking CRY-m in the dataset, BMAL TAD is lost as well (Figure 24). Interestingly, the BMAL TAD is absent in a number of species as shown in Table 31. These either miss half the TAD as shown in Figure 23 C or no BMAL (ARNT-like) is found but only ARNT. The assignment of ARNT is based on the distinct C-termini of ARNT, ARNTL1/BMAL1 and ARNTL2/BMAL2 homologs as shown in Figure

23 C. Either these species live without a BMAL or more probable in my opinion, the sequencing, sequence assembly or annotation are of insufficient quality. In the case of the partial TAD (*Bombylius major*) the sequence seems to be correct, or a sequencing error or wrong base call led to the interpretation of a stop codon less than 90 nt before the canonical stop codon would be found. Since most binding regions including the IxxLL harboring α-helix of the TAD are lost, it would most likely be dysfunctional. *Bombylius major* therefore probably has a unique, unfunctional TAD, that misses most of the interaction site.

Generally, it is a non-trivial task correctly assembling and annotating and here finding the correct sequence. ARNT and BMAL are very similar, possibly leading to BMAL sequences being assigned to an ARNT contig. If a *Drosophila* genome/transcriptome/proteome is used for search and annotation, the problem is even worse due to the differences between BMAL and CYCLE. If this is combined with low quality data, the reads for BMAL TAD might not be assigned correctly. The search for BMAL sequences will often return many CYCLE-like result. Not necessarily because there is no TAD but instead because the long unconserved region between the folded domains (bHLH, PAS-A, PAS-B) and TAD is ~ 150 aa in length (or 1/3 to 1/4 of the entire protein) will massively lower the alignment score. But careful examination of the sequences, and searches in TSA data allow to retrieve most data.

Table 31: Species with CRY-m but no BMAL TAD

Phylum	Class	Order	Suborder	Species	TaxID	comment
Arthropoda	Arachnida	Trombidiformes	Prostigmata	Tetranychus urticae	32264	Looks like ARNT
	Insecta	Coleoptera	Adephaga	Pogonus chalceus	235516	Looks like ARNT
		Diptera	Brachycera	Bombylius major	240869	Misses half the TAD

4.3.4.3 CLOCK PAS-B HI loop tryptophan is lost in absence of CRY-m

An important interaction between CLOCK and CRY-m is based on the PAS-B HI loop tryptophan thought to insert into the CRY-m secondary pocket and backed by mutational studies[35,84,92,96,103], yet so far no structural data for the interaction was presented. The same tryptophan is also responsible for PAS-B hetero and homodimerization in the CLOCK:BMAL and PER dimers.[35,93,95,101].

The tryptophan is replaced by a phenylalanine in all three brachycera species that lost CRY-m.

4.3.5 Interaction sites conservation

4.3.5.1 CRY-m:PER CBD

The CRY-m binding domain is highly conserved but to different extent over its length as shown in Figure 25. The N-terminal part shows higher conservation than the C-terminal part. Within the CBD the Chordata has a Kq insertion located the at the end of PER2 α3. Towards the end of the CBD an LSx insertion is found for insects at the end of PER2 α5. Both insertions are close to the end of the CRY-m CC helix[98] interaction site. PER diversifies at the C-terminus and is serine/threonine rich (hydroxyl amino acids) in insects where Chordata contain the CxxC motif. About 30 amino acids downstream the CBD, a region is found with the consensus sequence LVMIFEENAP. This region is conserved for all species in the study, but Chelicerata. It is followed by more prolines in insects, while in Chordata the glycine rich region proposed to be responsible for PER1 tail/BMAL TAD competition follows. This PER1 competing region shows similarities between all PER1 homologs.

4.3.5.2 CLOCK exon 19

The role of CLOCK exon19 (and its homologous sequences) is not entirely clear yet. The interaction with CIPC is well described but wasn't conclusively shown for the *Drosophila* homolog[79]. Exon19 is further described to be a transactivation domain (TAD) and able to recruit MLL1[76] in the mouse or TRITHORAX (TRX) in the monarch butterfly[78]. TRX is a homolog of MLL1, a histone methyltransferase important in transcription activation[76]. TRX catalytic activity was necessary for interaction in the monarch butterfly[78].

Most species in our analysis have an exon19 homologous region as shown in Figure 24 indicating conservation and functional importance. On the other hand, exon19 is missing in some species including the honeybee *Apis mellifera* and octopus *Octopus vulgaris* and the cuttlefish *Sepia esculenta*. How bees, octopus and cuttlefish compensate for the potential loss of MLL1 recruitment is unclear even more since their CLOCK proteins end before the exon19 site and don't seem to harbor unique, potentially compensatory features.

A sequence logo[167,168] of the exon19-like sequences in this study is shown in Figure 40 E. It shows all the main features derived from the crystal structure in complex with CIPC highlighted in Hou (2017) [79]:

Lxx**QL**xx**R**xxx**I**xxx**I**xx**QQ**x**EL**xx**I**xxx**L**

Additionally, we find:

(Lower case letters indicate lower conservation, brackets mean either or.)

(L/I)q(D/E)QL(Q/E)R(K/R)heeLQqlI(V/L/M)xQQEELR(R/K)(V/I)sEQLlm(A/V)rygll

that some positions between the red (from [79]) and bold (new) labeled ones are also well conserved either for identity, geometry (Q/E, very similar size, uncharged vs negative charge), chemistry (V/L/M,

hydrophobic) or charge (R/K or D/E). CIPC cannot be found in all species with an exon19 and exon19 cannot be found in all species with CIPC indicating multiple interaction partners. The interfaces forming the antiparallel coiled-coil bundle are therefore conserved in our species set. AlphaFold multimer was unable to predict anything useful for the potential interaction of CLOCK exon19 with the MLL1 core complex.

4.3.5.3 CLOCK polyQ

A polyQ motif in the *Drosophila* CLOCK C-terminus was proposed to work as a transactivation domain compensating for the loss of the BMAL TAD. While all three species without CRY-m have a CLOCK polyQ motif, some other species with an ancestral (n=4) or Hymenoptera (n=10) network structure also have a polyQ motif, including the mosquito *Anopheles gambiae*, fly *Bombylius major*, all mammals, reptiles, birds and fish and the sea urchin. The lack of conservation suggests convergent evolution instead of ancestry. PolyQ motifs can develop rather easily by expansion, a driver of Huntingtons disease[89]. The fast expansion could also mean that polyQ motifs are not selected but just the result of drift.

4.3.5.4 CRY-m disulfide and CRY-m CH:PER CxxC:Zn^{2+}

A Zn^{2+} interface between mouse CRY1 and PER2 Cry binding domain (CBD) was unexpectedly discovered by Schmalen et al (2014)[98]. CRY1 provides a cysteine (C414) and histidine (H473) residue while PER2 has a canonical CxxC motif (C1210, C1213) at the end of the CBD. Zn^{2+} binding was described to be mutually exclusive with a CRY1 intrachain disulfide bond between C363 and C412. CRY1 will be present in the oxidized form in the cytosol (disulfide bridge) and upon reduction will gain flexibility, enhancing complex formation which is stabilized by Zn^{2+} [98]. Nangle et al (2014) suggest in the same line of argumentation that the Zn^{2+} interface might have evolved to stabilize a flexible region[99].

I wondered if the disulfide bridge (yellow) and Zn interface (dark grey for CRY-m and PER) developed independently from each other (Figure 24). And indeed, the residues necessary for the disulfide bond are present in almost all species, indicating a conserved mechanism. Interestingly, they are missing in the Chelicerata (spider, mites, horseshoe crab). The latter cysteine is replaced by either leucine or isoleucine. Intriguingly, the Zn^{2+} interface residues C414 and H473 are conserved in all species studied even though only the PER proteins in the Chordata lineage have the corresponding CxxC motif. In this lineage CRY-m and PER got duplicated even multiple times resulting in up to 3 and 4 paralogues of PER and CRY-m, respectively. And all these PERs have the CxxC motif. Perhaps a much older metal binding site was repurposed in CRY-m for stabilization of the repressive complex. But only after gene duplication the PER proteins “learned” that trick (got selected for it).

The disulfide bond cysteine pair is not conserved in photolyases (checked for *Saccharomyces cerevisiae* and *Danio rerio* photolyases). It is also absent in *Drosophila* CRY-d but present in *Platynereis dumerilii*. The Zn^{2+} binding residues C414 and H473 are absent in both CRY-d and PL.

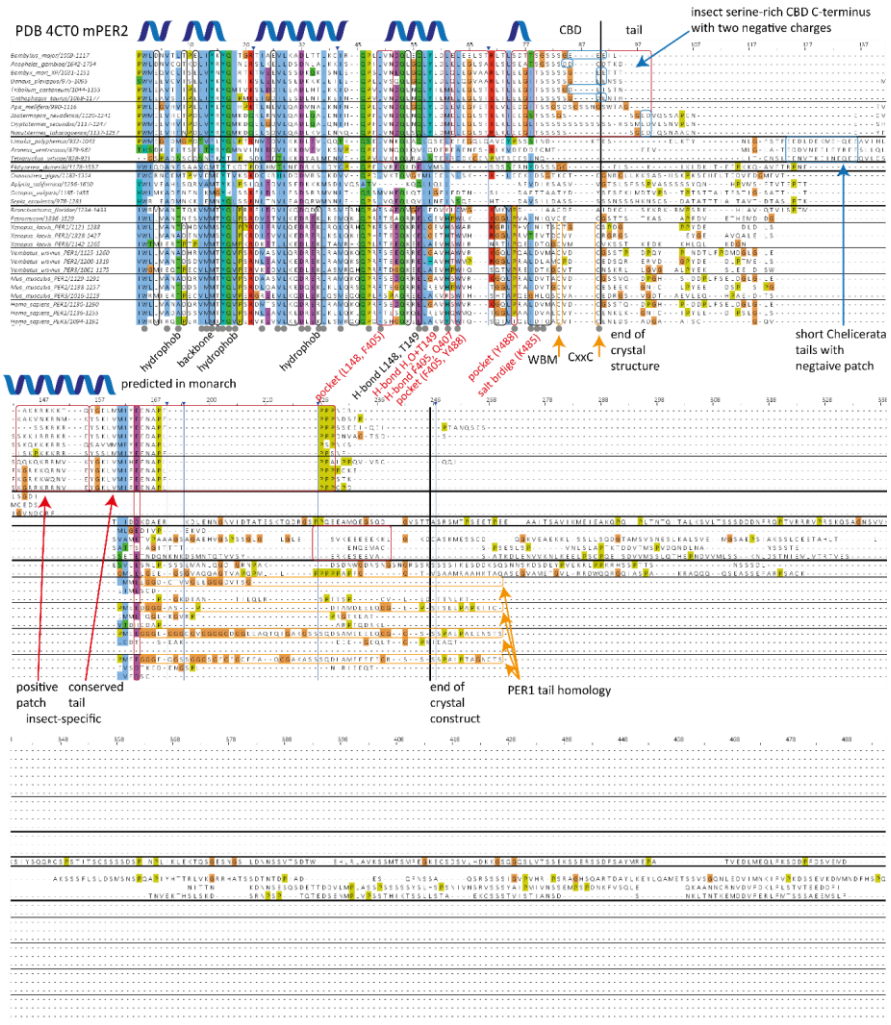

Figure 25: Multiple sequence alignment of the PER CBD and C-terminus. MSA starts starting from the crystal structure resolved CBD N-terminus. Thick lines indicate different taxa. From top to bottom: insects, Chelicerata (spider, horseshoe crab, mite), Annelida, Mollusca, Chordata. Thin lines separate subsocial from eusocial insects and chordates with multiple PERs from those with just one and also chordates with multiple PERs to better appreciate the paralogs. Blue lines indicate hidden columns. Columns containing only sequence from 1-5 species were hidden. Grey dots indicate mouse PER2 interacting residues according to the CRY1:PER2 crystal structure (PDB 4CT0)[98] *from which also the secondary structure assignment was used. For positions with low conservation the type of interaction*

is written below. Red font indicates a difference in the interaction from PER2(α4) to the Zn^{2+} interface. All CRY-m residues are conserved (Figure 17). A reduced species set was used to yield a readable figure. The PER tail shows taxa specific features such as high conservation in insects and exceptionally length in annelids and Mollusca.

4.3.5.5 PER CK1 BD

The casein kinase 1 binding domain (CK1 BD) on PER is very important because the recruited kinase is responsible for phosphorylation of not only PER but also of PER associated proteins[39] such as CRY-m and CLOCK resulting in their degradation or detachment from DNA[114]. Generally post-translation modification (PTM), with a focus on phosphorylation, comes increasingly into the focus in chronobiology also due to evolutionary relatedness even beyond conserved proteins[133–136].

Two CK1 BDs were defined recently[139] but the second one (CK1 BD-B) is more often described. They were not found in *Trichocorixia calva* (Water boatman) and *Heteropsilopus ingenuus* (flies) for both we only found partial sequences with sequence identity to the PER PAS domain core, but in all other species under investigation. This level of conservation emphasizes the functional importance of CK1 recruitment by PER. A sequence logo is shown in (Figure 40 B, D). The two amino acid gap in the middle stems from one *Ephemera danica* fragment. The CK1 BD is described to be α-helical and to form a stable interaction with PER[39,137,139].

4.3.5.6 PER WIN WBM

The interaction sites for WDR5, an important component of the MLL1 complex, are only present in PER proteins of *per* and *cry-m* gene duplicated Chordata species. and could have been acquired before gene duplication as was probably the case for the PER CxxC motif.

4.3.5.7 PER tail

The PER tail is defined as the sequence after the CRY binding domain (CBD). As discussed in the previous chapter, the competition of mouse PER1 with mouse BMAL1 TAD for CRY1/2 binding can be attributed to tail length and composition. The PER tails vary immensely in length as shown in Figure 27. The paralogue called PER1 always has the longest tail. Interestingly marine organisms seem to have a bias towards long tails. A high sequence similarity in insect PER tails is found.

4.3.5.8 CRY-m tail

The sequences after the photolyase homology region (PHR) are called tails. They are predicted to be unfolded and their physiological relevance is unclear but recent research highlights a role in delayed sleep phase syndrome (DSPS)[57,192]. A part of the tail binds to the PHR resulting in auto inhibition[62]. In mammals, the CRY2 paralog is shorter than the CRY paralog. In Lepidoptera (butterflies and moths) and

flies the CRY-m tails are extraordinarily long. The *Anopheles* tail is about as long as the PHR as shown in Figure 27.

4.3.5.9 CRY-m polyQ

PolyQ motifs were spotted in CRY-m tails, most notably in *Anopheles gambiae* CRY2. Because of the discussion on polyQ motifs in CLOCKs working as TADs, the same was analyzed for CRY-m. In insects only *Anopheles*, *Culex* and to a lower extent *Bombylius* have notable stretches (>5-10) of glutamine. *S. purpuratus*, the sea urchin, is the only other CRY-m with a Q rich tail. The functional relevance of these striking features is unclear, but the stretches are predicted by AlphaFold to be helical. If these polyQ tracts work in transcription regulation and why they sit in a repressor protein remains unsolved. Interestingly the long *Anopheles* tail which harbors the longest polyQ tract reduced CRY-m:BMAL affinity 2-fold, indicating an autoinhibitory function of long and/or polyQ tails in CRY-m.

4.3.5.10 FASP and per^s

CK1 phosphorylates [S/T] xx[S/T] motifs, where phosphorylation of the first hydroxylated amino acids enhances phosphorylation of the next. In mammals four of these motifs are continuous and mutation of the first serine leads to familial advanced sleep phase (FASP)[58,59,139]. The region is located between CK1BD-A and CK1BD-B and termed the FASP region. While the conservation within mammals is high, in insects the number of potential phosphorylation sites differs, and multiple stretches of repeats are present. Table 32 shows the number of phosphorylation sites and continuity per insect order. No continuous phosphorylation sites are found in Hymenoptera. In lower Diptera (mosquitos) up to 3 continuous sites are present. Continuity improves phosphorylation speed because every phosphorylation at position -3 results in efficient recruitment of CK1 and thereby catalyzes phosphorylation[139131,139]. Sequential phosphorylation of PER by tandem phosphorylation sites does not seem to be conserved in insects.

Table 32: Phospho-sites in the FASP region by insect order. Alignments of the potential phosphorylation sites are shown in Figure 41.

Clade	Phosphorylation sites X..X	Type
Chordata	5	Continuous X..X..X..X..X
Hymenoptera	discontinuous	discontinuous
Lepidoptera	Up to 3 continuous	
Diptera (low)	Up to 3 continuous	X..X..X,
Diptera (high)	2-3 continuous	X..X, X..X..X
Coleoptera	Up to 4 continuous	X..X..X..X

The *Drosophila* per^{s} (per short) mutant shows shortened circadian phase. The per^{s} mutant carries a mutation at S589 and is followed by the per^{sd} region (per short downstream 604-629)[140]. Similarly, to the FASP region per^{s} and per^{sd} control the stability of PER and lie just before the first CK1 binding site. The per^{s} and per^{sd} regions are much better conserved in insects than the FASP as shown in Figure 26.

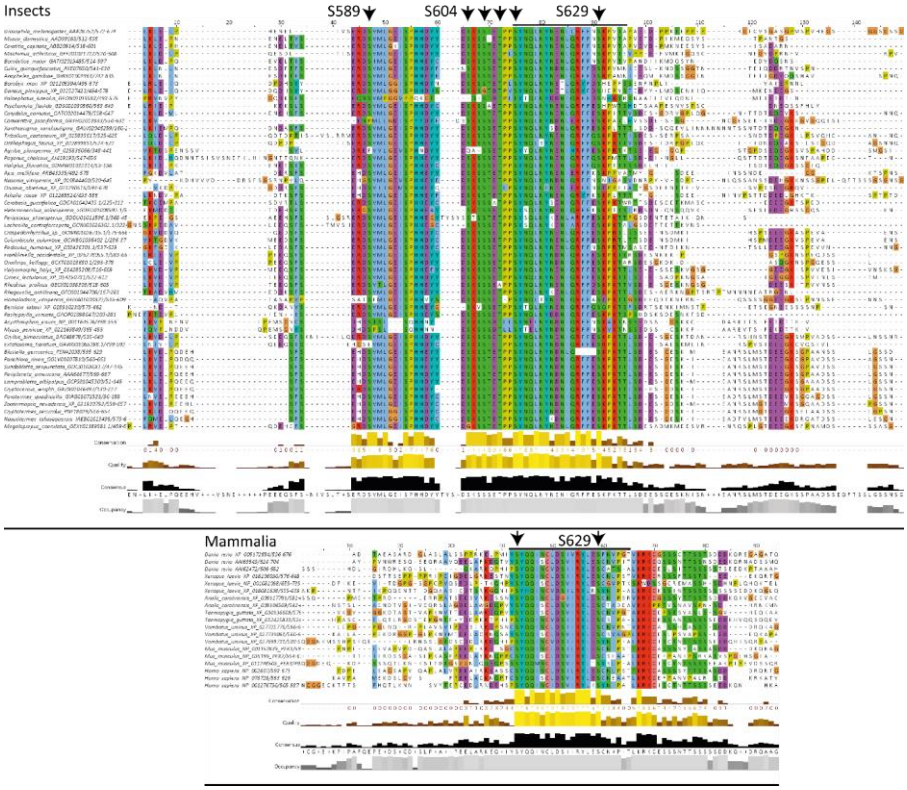

Figure 26: Conservation of per^{s} site. The CK1 BD is indicated by the black line, phosphorylation sites by arrow heads. Insect and mammals were split in two groups to strengthen the conservation seen in each group. The CK1 BD is conserved but has lineage specific differences.

4.3.6 Summary

The small number of marine organisms in this study, namely Hemichordata (marine acorn worms) and Echinodermata (sea stars, sea urchins) might have lost PER. How they sustain circadian rhythms and what importance circadian rhythms have is unclear. We can further appreciate the high level of conservation of the proteins and interactions in the central loop across time, phyla, life styles, biological rhythms such as diurnal/nocturnal, seasonality, sizes and environments. Some interesting observation can be made for marine organisms and Chelicerata tough. Their PERs are distinct in length as shown in

Figure 27 and composition as shown in Figure 25, respectively. Even more surprising since the marine organism are not necessary the closest relatives within this study, an argument for convergent evolution.

We showed that not only is the circadian clock network structure of the monarch butterfly ancestral, even though the Lepidoptera are the second youngest insect order, but as old as the insect class and even older since we have to expect that the last common ancestor of arthropods and Mollusca and also Chordata must have had such a network more than 700 million years ago. We saw that the network structure can be identified from the presence of CRY types present, but also from the presence of a BMAL TAD or if good data is available from PER CBD presence. Using all three indicators increases certainty. Network structure coincides sometimes with insect order or suborder. How and why the brachycera network structure evolved will continue to puzzle me.

We saw that the CK1 binding domains on PER are preserved but the FASP phosphorylation site is not. This indicates that the FASP sites is also an innovation of the CRY-m/PER gene duplicating Chordata such as Zn^{2+} binding site and WBM and WIN motifs.

The CRY1 tail exon 12 appears to interact with PER. Both, the insect PER tail and chordate PER1 tails show conservation with is functionally relevant.

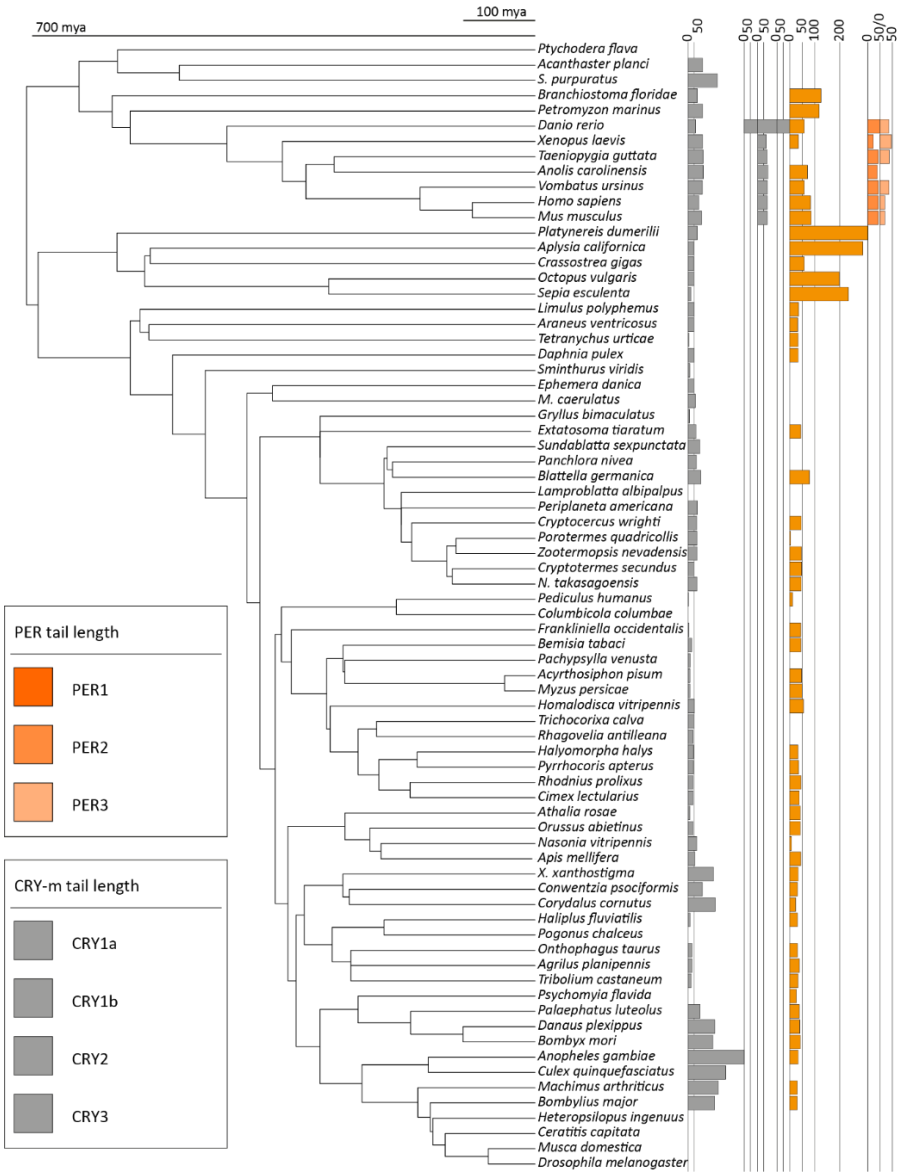

Figure 27: Lengths of CRY-m and PER tails. CRY-m tails are shown in grey; paralogue annotation was combined. PER tails are in shades of orange. Anopheles CRY-m tail is about as long as the PHR. PER1 tails are the longest of the paralogue, with a bias to longer tails in marine organisms. Heteropsilopus ingenuus and Ptychodera flava CRY-m sequences are incomplete and tail lengths couldn't be determined.

5 Discussion

5.1 The CRY-m:BMAL1 TAD interaction

5.1.1 Critical literature review

The interaction of the repressor CRY-m with the activator BMAL is described and accepted in the literature to be between the last helix (CC-helix) of the CRY-m photolyase homology region and the BMAL C-terminal transactivation domain (TAD)[71,83]. I challenge this view and propose that the BMAL TAD indeed binds to the ancestral FAD binding pocket also called the primary pocket.

Chaves et al (2006) were the first to suggest that the CRY1 CC-helix is important for repression. This was based on the (i) inability to CoIP PER1/2 by CRY1 ΔCC and CRY1 ΔCC-tail, (ii) reduced nuclear localization of PER2 with CRY1 ΔCC, (iii) reduced and basically lost CLOCK:BMAL reporter gene repression by CRY1 ΔCC and CRY1 ΔCC-tail, respectively and (iv) the ability of a chimeric *Arabidopsis thaliana* (6-4) PL with an extended mouse CRY1 CC-tail (371-606) to repress[97].

Secondly, Czarna, Breitkreuz et al (2011) showed that CRY1 and CRY2 CC-helix + tail (CCT) peptides interact with BMAL1 TAD proteins (577-626 and 490-626) by fluorescence polarization and ITC. The ITC showed endothermic (heat is absorbed) binding events that are entropically favored. The authors further performed a peptide scan analysis in which $CRY1_{471-606}$ and $CRY2_{489-592}$ representing peptides were immobilized on nitrocellulose and incubated with fluorescently labeled $BMAL1_{577-625}$ protein. The overlapping 10-mer peptides showing binding represented $CRY1_{473-504}$ and $CRY1_{565-595}$. Alanine substitutions of negatively charged amino acids increased $BMAL1_{577-625}$ binding while the reverse was the case for substitutions of basic amino acids, suggesting that positive charges were necessary for binding[83].

Xu, Gustafson et al (2015) also used isolated CRY1 peptides, here only representing the CRY1-CC-helix ($mCRY1_{471-505}$). They measured affinities by fluorescence polarization and ITC for BMAL1 TAD mutants and observed a correlation for $BMAL1_{579-626}$:CRY1 CC affinity to period length (lower K_D – shorter period). Their ITC also showed an endothermic reaction. Additionally, NMR chemical shifts were assigned identifying the binding site to be on the α-helical and C-terminal parts of the BMAL1 TAD and demonstrated a spatial proximity of the two induced by CRY1 CC. They further showed an importance of the BMAL1 TAD C-terminus as a 619X mutants lacking the seven last amino acids neither bound CRY1 CC-helix in ITC, nor produced strong amplitudes[71].

The presented biochemical binding data can be well explained with a coiled-coil interaction in solution between the CRY-m CC-helix and BMAL TAD α-helix in which the C-terminus folds back onto the heterodimer interface as shown in Figure 28 B. 2013, Czarna et al presented the structure of apo CRY1

showing that the CC-helix is in contact with the PHR and not sticking into the solvent[101]. About 70% of the helix are solvent-accessible, so the interaction could still happen as in solution if the necessary helical interface is presented to the solvent. 2014, Schmalen et al and Nangle and Rosensweig et al published structures of CRY1:PER2 and CRY2:PER2 respectively. Both publications show that PER2 is wrapping tightly around the CRY CC-helix[98,99]. Yet, Fribourgh, Srivastava et al (2020) present fluorescence polarization data that shows insignificant differences in binding of a fluorescently labeled BMAL1 TAD peptide to either apo CRY1/2 PHR or CRY1/2 PHR:PER2 complexes[96]. Importantly, the CRY-m:PER complexes were preformed and the CRY-m:PER CBD complex has a higher affinity (~10 nM, see Figure 38) than any affinity reported for CRY-m:BMAL TAD in the publications above (0.8-10 µM). A lack of binding competition between an at least 100-fold stronger binder, that bound first, and a weaker binder is highly unlikely. Further, the binding data by Fribourgh, Srivastava et al appears more relevant than the data presented by by Xu, Gustafson et al (2015)[71] and Czarna, Breitkreuz et al (2011)[83] because the CRY-m PHR is used instead of a peptide in an unphysiological state (solution instead of bound to PHR). The usage of a BMAL1 TAD peptide seems more logical to me as the motif is predicted to be unfolded and this is supported by CD-spectroscopic measurements by Czarna, Breitkreuz et al (2011)[83]. Further I want to mention that the publications of Czarna, Breitkreuz et al resulted from our lab, the criticism does not go against competing labs or highly accomplished former lab members. I rather want to challenge a model that established in the chronobiological community based on conclusions that at their time seemed logical and correct and which appear questionable more than 14 years later.

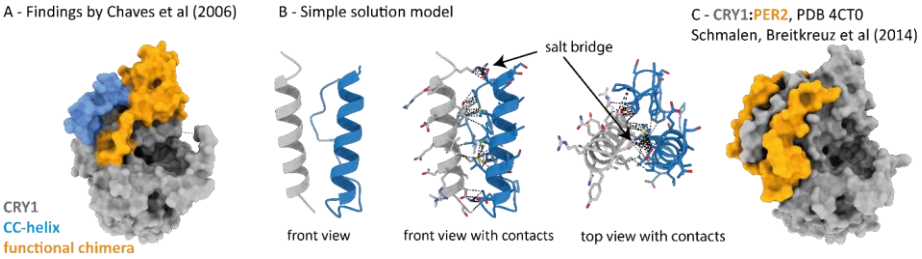

Figure 28: Simplistic model for CRY CC:BMAL TAD interactions. A: The ΔCC mutants by Chaves et al were lacking the blue colored CC-helix. The same is tightly enclosed by PER2 in C. Arabidopisis (6-4) PL was rendered a repressor by including $CRY1_{371-606}$ (in blue + orange) into a chimera. B: An AlphaFold based model for the interaction of $CRY1_{471-496}$ CC (dark grey) and $BMAL1_{595-626}$ TAD (blue). Black dashed lines indicate contacts. A salt bridge and a well packed hydrophobic core are shown. C: The PER CBD wraps around the CRY-m CC helix, covering most of the solvent accessible surface.

Interestingly the necessity for the longer fragment $CRY1_{371-606}$ to generate a repressing chimera was ignored. The correlation between CRY1 CC helix:BMAL1 TAD affinities and period length cannot easily be explained by the solution model. I can only speculate that the apparent correlation is either an artifact or randomly fits. Of course, the possibility cannot be excluded that the BMAL TAD binds to the CRY-m CC-helix in absence of PER and solution data can partially recapitulate the domain integrated state. If my assessment of the literature is correct little information is available hinting to a binding site: (i) located somewhere on the CRY-m PHR, but not the tail and (ii) that $CRY1_{371-606}$ is necessary and sufficient in the context of a (6-4) PL chimera to result in efficient repression. The effect of several amino acid substitutions in the BMAL1 TAD were characterized by oscillation experiments[71]. A putative binding site should explain these.

5.1.2 Where does BMAL1 TAD bind to CRY?

In the absence of experimental data, structure prediction was used to get a first idea of the binding site. While structure prediction cannot give evidence of an interaction it does provide a model which can be tested experimentally and is therefore great to build a working hypothesis. Here AlphaFold was used because it is the best scoring prediction system in the critical assessment of protein structure prediction (CASP). For a longer discussion on the value of structure prediction and especially AlphaFold see the discussion section on AlphaFold.

A benchmarked methodology for domain-motif interactions was employed because the prediction score was low and based on anecdotal evidence AlphaFold likes to stick small peptides into pockets. This approach uses peptides of varying length with and without the motif and is scored based on the fraction of common contacts/recurrent spatial closeness. Only distance is used for evaluation after the model is build which is intrinsically scored by AlphaFold. AlphaFold reproducibly placed the BMAL1 TAD in the CRY1 primary pocket as shown in Figure 14. The primary pocket is lined by the protrusion, phosphate binding and C-terminal lid, which all show CRY-m specific conservation compared to CRY-d (Figure 8). The CRY1 primary pocket is lined with positive charges and is the binding site of negatively charged FAD in related molecules. The orientation of the negatively charged BMAL1 TAD was predicted in multiple ways which all satisfy the contacts between hydrophobic and hydrophilic amino acids. The apparent flexibility in binding would explain why no structural data was presented so far of the interaction even though several people must have tried to do so considering the importance of the binding site, the many years of knowledge and availability of material. Interestingly, the recent advances in cryo electron microscopy might allow to at least see density of the BMAL1 TAD on CRY1 but is quite limited by the relatively small size of the complex. The binding site would also explain the ITC data from Czarna et al (2013)[101] with full-length CRY1 and BMAL1 fragments that showed an endothermic (entropy-controlled) reaction. Ordered waters on the relatively hydrophobic BMAL TAD

would be freed by the interaction increasing entropy. Due to the high flexibility of the model this should not be over interpreted. Only molecular dynamics simulations on these models with the mutants would potentially allow a useful statement.

The binding site was further investigated using a crosslinking approach. The BMAL1 TAD was used in the form of a peptide which seems like a safe option because (i) this was done previously for the interaction with CRY-m PHR[96] and (ii) the lack of native tertiary structure as measured by CD spectroscopy of *E. coli* expressed larger BMAL1 fragments ($BMAL1_{477-626}$, $BMAL1_{530-626}$)[83]. Additionally, corresponding fragments from the monarch butterfly BMAL didn't yield soluble protein but peptides are soluble, and a comparison was the initial aim. A photo-crosslinker moiety was synthesized onto the N-terminus of the BMAL1 TAD peptide by a collaborator. A trifluoromethyl phenyl diazirine was chosen because diazirines only require UV light at ~360 nm which was available and more importantly is not absorbed by protein. Excited diazirine yields a carbene after nitrogen loss ($-N_2$) which reduces unspecific crosslinks due to its short-lived nature while being able to react with every kind of amino acid in its vicinity due to its radical nature.

The results in Figure 14 show crosslinks of the BMAL1 TAD at the primary pocket of CRY1 and the typically unresolved phosphate binding loop as well as some on the CRY1 CC-helix. The phosphate binding loop closes the primary pocket in CRY-d, binding and capturing FAD. Crosslinking to this loop could indicate a similar closing of the CRY-m primary pocket after BMAL TAD association or just binding. These crosslinks were found in at least three out of four CRY-m proteins: mouse CRY1 PHR, CRY1 full-length, CRY2 PHR and CRY2 full-length. The overlap between samples is good. Crosslinks are only expected close to the BMAL1 TAD N-terminus. The positions can be explained by binding of the BMAL1 TAD into the CRY-m primary pocket with sufficient mobility within the pocket. Alternatively, multiple binding sites have to be assumed. This data is hinting towards involvement of other parts than the CRY-m CC-helix.

Binding assays indicate importance of the primary pocket for TAD binding. Validation of structural models by biochemical and cell biological assays is very typical to test the model and the importance of found interactions. Three of the designed CRY1 mutants (phosphate loop R236A+R238A, pocket primary H355A, C-terminal lid Q407E+Q408E) reduced binding in the fluorescence polarization binding assay. Importantly, the reducing mutants are located on opposite sides of the primary pocket. The same mutants either decreased (R236A+R238A) or increased repression (H355A, Q407E+Q408E) of BMAL1:CLOCK activity by CRY1 as shown in Figure 15. The cellular system is a more complex one and already the inputs are different. The interaction of CRY1 and CLOCK can have an impact, but the used mutants are far from the proposed binding site at the secondary pocket. The influence of CLOCK:CRY1 binding can therefore be assumed to be constant. Loss of interaction with the E3 ubiquitin ligases FBXL3

is demonstrated to lead to phase delay due to increased CRY1 levels and prolonged repression[106]. Here repression is measured in a steady-state system, but the protein concentrations are not as controlled as in the biochemical binding assay. Reduced FBXL3-dependent degradation, i. e. enhanced CRY1 stability can explain the increased BMAL1/CLOCK repression by the CRY1 mutants (H355A, Q407E+Q408E, H411A, R421A) which are part of the FBXL3 binding site as explained in the Results section.

The presented data hints towards involvement of the CRY1 primary pocket and especially the phosphate binding loop for BMAL1 TAD binding but is not conclusive. Further, no mutants of the CRY1 CC-helix were designed and tested to verify or falsify the literature. The CRY1 K485D mutant, located in the CC-helix, does not impact CRY1 repression of BMAL1:CLOCK in reporter gene assays. But K485D/E also reduces PER and FBXL3 binding in fluorescence two hybrid and luciferase complementation assays[101]. This mutant reduces importance of the CRY1 CC-helix in BMAL1 TAD binding. Yet, in the same line of argumentation, its reduced interaction with FBXL3 could also result in higher cellular levels masking a reduced repression in the reporter gene assay. The same applies to the CRY1 E382R+E383R mutant from the same assay. A third mutant, the CRY1 G336D mutant is unable to repress BMAL1:CLOCK, but is also unable to bind to PER2. Both effects can be explained by a disruption of the structure that the available crystal structure suggests for this position.

It was overlooked to include CRY1 CC-helix mutations when mutants were planned because the mutants were based on crosslinking, AlphaFold and positive charges. The mutant approach lacks the throughput to efficiently test many mutants. There are three bottlenecks in the chosen approach:

1. Generating site specific mutants is time and cost expensive. Mutants have to be planned and ideally checked before for possible interactions with known binding sites. Primers have to be purchased, mutagenesis has to work, and clones analyzed. Here this had to be done for the reporter gene assay and the binding assay in different plasmids, doubling effort.
2. For the binding assay, purified protein is required. Expressing cryptochromes in *E. coli* doesn't yield well folded and functional protein instead expression in insect cells is necessary for biochemical assays.
3. Insect cell expression requires an at least two-week protocol up front which can only be parallelized to a certain degree: bacmid has to be generated, isolated and transfected. The resulting baculovirus has to be amplified and tested. The protein has to be expressed in insect culture. Finally, protein has to be purified in multiple steps.

This results in the inability to test many mutants in an unbiased, cost and time efficient way. This is the reason why not more mutants were tested and why mutants were not combined. Combination of

mutants is probably required for larger effect sizes. The assumed electrostatic interaction for the interaction must result in low effects by single or double alanine mutants as observed.

A different approach could yield more meaningful results faster. The expression system of choice would be yeast. Yeast expression is often able to yield functional mammalian proteins since it is also an eukaryote and contains some of the necessary protein folding machinery for folding of mammalian proteins. Yeast grows faster than insect or mammalian cell culture and transformation is sufficient for DNA delivery. Furthermore, many assays for protein-protein interaction (PPI) are established in yeast. Most PPI assays aim at finding interactors from large libraries or improve binding like in surface or ribosome display methods. Here we are challenged with the reverse problem, an interaction for which we want to find interaction reducing mutations. The interaction of CRY-d with PER and TIM-d was previously tested by yeast two hybrid assay (Y2H) [201], showing that CRYs can be functionally active expressed in yeast and in this assay. Y2H was originally developed to screen libraries for PPIs but can be used to screen for binding sites. In Y2H one partner, the bait is fused to a Gal4 DNA binding domain. Reporter genes in the yeast have Gal4 binding sites in their promoters. The second partner, the prey protein is fused to the Gal4 activation domain. Bait and prey interaction results in reporter gene expression allowing growth on histidine and adenine deficient medium and α- and β-galactosidase activity. Screening would be performed by growth analysis of serial dilution of transformants on selection media and β-galactosidase activity. Many mutants can be generated in an unbiased approach by random mutagenesis for example by mutagenic PCR which includes manganese (Mn^{2+}). This method is very fast and easy. It doesn't fully explore the possible diversity since single nucleotide exchanges in a codon don't allow to get to every other amino acid. Instead, DNA shuffling using CRY-d genes would be an option to create a diverse library since *Drosophila* CRY-d doesn't bind to CLOCK/BMAL in EMSA[187]. But other approaches including DNA shuffling or purchasing random libraries are more complex, expensive, and time-consuming. The Gateway cloning system allows to clone generated CRY1 mutants into the plasmid for Y2H which would be sequenced to exclude non-sense mutants. Known mutants would then be transformed into yeast and screened. Growth impaired and β-galactosidase inactive mutants would be considered to have impaired BMAL1 TAD:CRY-m interaction. These mutants could then be tested in a reporter gene assay for validation. The reporter gene assay also uses Gateway plasmids enabling fast shuffling of the mutant genes without the need to reclone. Limiting the number of mutants to ~100 and knowing their identity would allow to deduce a binding site.

There are many CRY-m specific conserved positions around the primary and secondary pocket as well as the α/β domain as shown in Figure 8. Since the BMAL TAD doesn't interact with CRY-d, these positions can be considered potential binding sites of CRY-m specific interactors. Space-wise the primary pocket or the interface between the primary pocket, C-terminal lid and the CC-helix which is

full of aromatic amino acids appears to be the best option to place the BMAL TAD. While AlphaFold predicts the helix to sit inside the primary pocket, the crosslinking data would support a model where the helix lays on top of the primary pocket with its N-terminus towards the phosphate binding loop.

5.1.3 Evolution of interaction

The conservation of the CRY-m:BMAL TAD interaction and even charge preference was demonstrated for a small number of cryptochromes. Originally an ant (*Temnothorax nylanderi*) and a beetle (*Tribolium castaneum*, red flour beetle) CRY-m should have been included to cover more of the well-studied insect orders. Unfortunately, no coding sequence (CDS) or sufficient protein could be amplified or purified, respectively. The data still shows preference for negatively charged TADs and this can, at least partially, explain the differences in affinity between mouse and monarch BMAL TAD. The usage of the small BMAL TAD C-terminus peptides can be considered problematic because they, according to the crosslinking, bind less specific than the full BMAL TAD as shown in Figure 14 B. Importantly, the peptides were used at 50 nM for the binding assays and much less than the protein (up to 100 µM) while for the crosslinking the peptides were use at 3-fold molar excess at 300 µM. These are very different conditions. Ideally only full BMAL TAD peptides would have been used. Synthesis of these long (>30 aa) and lowly charged peptides is not always possible. The synthesis of the monarch BMAL TAD peptide took multiple attempts by the producer. Also, labeling and following purification of the peptides by the collaboration partner on campus, Dr. Fabian Barthels, is not straightforward with the longer peptides and already had to be optimized for the smaller uncharged peptides. The evolutionary perspective first motivated to investigate BMAL TAD sequences and allowed to find the charge dependency which seems to be most pronounced for mammalian CRY-m. This could imply an increasing selection for the negative charges present in mammalian and most TADs as shown in Figure 16 A. It further seems like the BMAL TAD in insects became less acidic and added a C-terminal proline, that does however not alter the binding affinity.

5.1.4 Activatior preference are more complex

The reporter gene assay nicely validated the binding assay. The repression within this assay was based on CRY1 only. The co-activator preferences are depicted in Figure 29. While CRY-m prefers a negative charged BMAL TAD C-terminus as shown by fluorescence polarization and reporter gene assays in Figure 16, co-activation prefers the D->N substitution (insect C-terminus (inC-term) to Lepidoptera C-terminus (lep C-term)) and the glycine instead of serine at the N-terminus of the TAD. Aspartate to glycine exchanges lower the activation, so negative charges at these positions seem to be relevant.

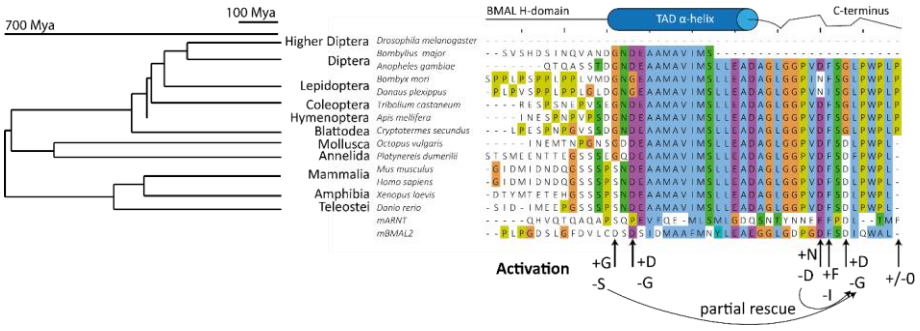

Figure 29: BMAL co-activator preferences. A S->G substitution partially rescues the effect of the single C-terminal D->G and D->N exchange. But the neighboring N-terminal D->G exchange reduces activation. The C-terminal proline does not affect activation.

5.2 Competitive binding of PERIOD and BMAL1 TAD to repressor CRY1

Mammalian PER paralogs are considered co-repressors. The literature focuses on PER2 granting it the bigger role while PER1 remains understudied even though it is considered important for the central loop. PERs are part of the early repressive complex and recruit CK1 to the activation complex. The importance of phosphorylation of CLOCK is better understood now and the evolutionary conservation of this repression mechanism is obvious. The exact temporal order of events is still not clear especially because studies with temporal resolution of clock mechanism are difficult to conduct. A landmark study in this regard was published by Koike et al (2012)[37] who established the model of the early and late repressive complex by showing that PERs and CRYs are present together from CT12 to CT20 and from CT0 to CT4 CRY1 alone resides at circadian promoters.

CRY-m is known to repress CLOCK:BMAL activity by binding of CLOCK W362 and BMAL TAD. The importance of the BMAL1 TAD for activation and competition of binding by CRY-m are established concepts. Additional repression by PER is considered a co-repression function. Many of these views are based on reporter gene assays working with steady-state conditions instead of temporal resolution. Fribourgh and Srivastava et al (2020)[96] show that PER2 doesn't impair the interaction of CRY1 PHR or CRY2 PHR with a fluorescently labeled BMAL1 TAD in a fluorescence polarization assay further supporting the idea of PER adding functions to repressive CRY-m without negatively impacting them.

5.2.1 Model of competition

Kommentiert [CM3]: Evtl neu machen, definitiv BildUnterschrift ranziehen

Yet, the presented data in the chapter "Competitive binding of PERIOD and BMAL1 TAD to repressor CRY" (from page 76) clearly shows that PER1 and PER2 CRY-binding domains (CBD) and C-terminal tails reduce CRY-m:BMAL1 TAD affinity, impact CRY-m:BMAL TAD crosslinking and BMAL:CLOCK repression under steady-state conditions. The relative influence of the PER CBD and tails were investigated and

show paralog specific competitive mechanisms in which the PER1 tail and PER2 CBD show the strongest impact. Additionally, the PER1 and PER2 "core" region, which includes all domains preceding the CBD and therefore the PAS-/PAS-B homo- and heterodimerization domains and FASP, phosphodegron and CK1 binding sites, affects BMAL1:CLOCK in a different manner.

While the PER1 core does not affect CRY-m mediated BMAL1:CLOCK repression, the PER2 core fulfills the described co-repressor function. Full-length PER1 however is able to lift BMAL1:CLOCK repression, and this effect is mostly mediated by its C-terminal tail region, that also weakens BMAL1 TAD:CRY-m affinity (Figure 22). According to the established role of PERs and described reduced binding of CK1 to PER1[39] PER1 may also lift CRY1:BMAL1 TAD-based repression by recruiting insufficient amounts of CK1 to the CLOCK:BMAL heterodimer. Other possibilities are different concentrations, PTMs, stability or a combination of them.

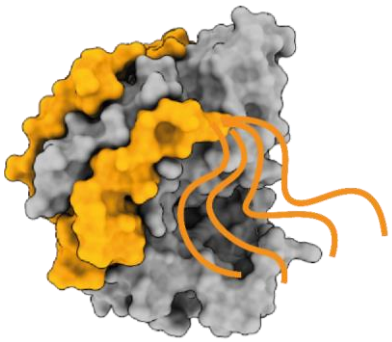

Figure 30: Model of PER1 tail (orange) repression.

PER1 CBD (orange) wraps around CRY1(grey) and the flexible tail blocks the primary pocket.

The PER1 tail is predicted to be intrinsically disordered, which is obvious by the glycine and proline richnessand well documented in the crosslinking data (Figure 20, Figure 21). The above model of binding of the BMAL1 TAD to the CRY-m primary pocket together with the proximity of the PER tail and PER1s disorder suggests that molecular crowding reduces BMAL1 TAD binding in presence of PER1. The flexibility and negative charge of the tail could partially occupy the CRY-m primary pocket and reduce interaction of the weaker binder due to high local concentrations as shown Figure 30.

The effect of PER1 CBD binding to CRY-m on CLOCK binding is unknown. The PER2 CBD negatively affects the interaction of CRY1:CLOCK (3-fold reduction) while increasing the interaction of CRY2:CLOCK (2-fold increase, Table 1)[96]. The above analysis didn't take into account the effect of PER CBDs on CRY1:CLOCK. Repeating the analysis by only comparing samples with the same CBD, yields the same result that the PER tail and PER2 CBD compete with the BMAL1 TAD for binding to CRY1. The effect of the PER2 CBD was explained by the PER CBD decreasing the flexibility of a serine loop close to the CLOCK binding site on CRY1 in molecular dynamics simulations. Similar simulations were not performed for PER1 and the effect of PER1 CBD remains unknown.

The effect of the PER2 CBD and to a lesser extent PER1 CBD on BMAL1 TAD binding is less clear. Structural information for the CRY-m:PER1 interaction is lacking and would be very informative for the CLOCK:CRY-m, BMAL TAD:CRY-m interaction and insect CRY-m:PER interaction.

The protection against FBXL3 binding and subsequent degradation was only demonstrated for PER2[42] but by homology should also work for PER1. If the model for BMAL1 TAD binding is correct also BMAL1 TAD presence should protect to a lower extent than PERs against FBXL3/21 because PERs cover a larger fraction of the FBXL3 interaction site and has a higher binding affinity.

5.2.2 Why was this effect not observed earlier?

Why didn't Fribourgh, Srivastav and Sandate et al (2020)[96] measure the competition between PER and BMAL for CRY-m binding? According to the method section they used human (h) $PER2_{1095-1215}$ CBD, while here mouse (m)$PER2_{1119-1215}$ CBD and $PER2_{1119-1252}$ CBD+tail were used. The main effect of PER2 on CRY-m:BMAL1 TAD interactions stems from the CBD so they should have observed a change in K_D. Their construct is 24 amino acids longer on the N-terminus but the herein presented reporter gene assays include full length protein and replicates the fluorescence polarization results. The human and mouse PER2 CBD fragments are 74% identical, an alignment is shown in Figure 31. They used a slightly different buffer (50 mM Bis-Tris Propane buffer pH 7.5, 100 mM NaCl, 2 mM TCEP and 0.05% (vol/vol) Tween-20), which is largely similar to the one used herein (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 0.05% Tween-20). Both buffers were tested and resulted in indistinguishable results. Their TAMRA-$hBMAL_{594-626}$ TAD probe was diluted to a working concentration of 50 nM in assay buffer, identical with the FAM-mBMAL $TAD_{600-632}$ probe used in this thesis besides the fluorophore. Crystal structures from both PER2 constructs with CRY1 exist (PDB 4CT0[98] and 6OF7[96]) and are also almost identical with a root mean square deviation (rmsd) of 0.688 which would be even good for two structures of the same protein. The difference in binding results remains mysterious.

Figure 31: Alignment of PER CBDs from human $hPER2_{1095-1212}$ (top, PDB 6OF7) and mouse $mPER2_{1119-1252}$ (bottom, PDB 4CT0). The sequence similarity is very high and all CRY1 contacting residues are conserved (Figure 31).

Interestingly there are multiple studies and suggestions in the literature showing a reduction of CRY-m binding to CLOCK:BMAL in the presence of PER2 without gaining general recognition. Ye et al (2011) showed reduced binding of hCRY1 to CLOCK:BMAL on E-box DNA in EMSAs by PER2[187]. Akashi et al (2014) showed dose-dependent increasing CLOCK:BMAL1 activity in presence of either PER1 or PER2 in reporter gene assays using a *per2*-Luc reporter. The $PER1_{1084-1291}$ CBD or $PER2_{1068-1257}$ CBD were necessary and $PER2_{1068-1257}$ CBD sufficient[202]. Rosensweig mentions these results 2014 (CRY2:PER2

structure)[99] and in his review 2018[2]. The result by Fribourgh and Srivastava et al (2020)[96] can be understood as a response to him mentioning a potential PER2/BMAL TAD competition since the paper is largely based on Rosensweig (2018) [103] where he studied the difference between CRY1 and CRY2 and also mentions the possibility of such a competition.

5.2.3 How does this change the model of the central loop?

The model of the central loop is updated by the presented results. In the early repressive phase CRY1/2 probably do not significantly reduce CBP/p300 recruitment to the BMAL TAD because of their association with PER1 and PER2. Latest at CLOCK hyperphosphorylation and detachment from DNA will this repression mechanism be unnecessary. If CLOCK:BMAL don't bind to chromatin, then the transcriptional regulators aren't recruited chromatin as well (DNA off mechanism of repression). Instead the competition of CBP/p300 and CRY1 gets relevant once PERs are degraded or at least not associated anymore with CLOCK:BMAL in the late repressive phase. CRY1 can now bind to the BMAL TAD and after CRY1 degradation the clock is reset. Interestingly, the CLOCK PAS-B W362, described to bind to the CRY-m secondary pocket, binds to histone 3 residues 75-82 in the recent cryo electron microscopy study by Michael and Stoos (2023)[36]. CRY-m association could therefore proceed as follows: PER-bound CRY-m bind to CLOCK W362 loosening CLOCK:BMAL from nucleosomes (DNA on repression). PER-associated CK1 phosphorylates CLOCK leading to hyperphosphorylation and dissociation from DNA (DNA off repression). Phosphorylation of PERs results in their degradation. CRY1 can now sequester the BMAL TAD resulting in loss of CBP/p300 association (late repressive complex) as shown in Figure 32.

This emphasizes the importance of the multistep process of repression. Full CLOCK:BMAL inactivation is only possible in a late repressive phase when PERs are degraded. CRY2 is only represent in ChIP seq when PER2 is present according to Koike et al (2012) [37] which implies that CRY2 is not able to bind BMAL TAD if not the high local concentrations and avidity by binding of the CLOCK W362 make an association viable again.

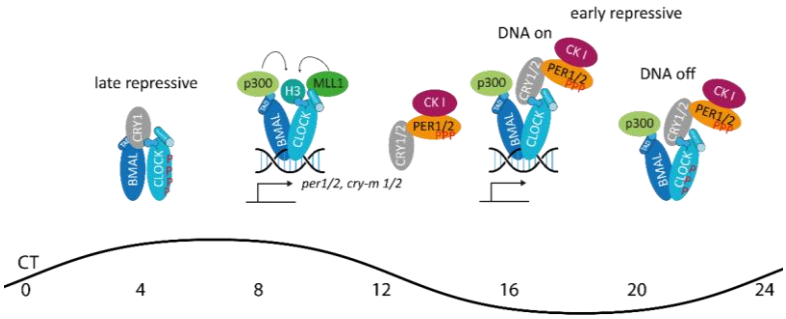

Figure 32: Updated model. This model excludes a CRY:BMAL TAD interaction in PER presence. Interpretation of reporter gene assays should include the BMAL/PER competition.

5.3 Evolutionary analysis of interactions sites in the circadian clock central loop

5.3.1 CRY interaction sites evolution/conversation

The circadian clock system is quite complex with many protein interactions, PTMs, transcription regulation, nuclear localization and so on. This is even more astonishing when the conservation of the system is considered which reveals that all animals inherited a 700 mya old system. This system arose with the first animals. The usage of cryptochromes in plant clocks could hint towards an even more ancestral rudimentary clock that the common ancestor of plants and animals used. Yet only CRYs ancestral photoreceptive mechanism is used while no other clock components is conserved suggesting rather convergent evolution.

5.3.2 Do we need to look for interaction sites or is gene presence a sufficient indicator?

The evolutionary analysis started to find out which insect clades rely on which circadian clock mechanism: *Drosophila*-like, mammalian/honey bee-like or ancestral/monarch-like. It seemed that *cry* genes have to be found and identified as transcription repressor *cry-m* or photoreceptor *cry-d*. This analysis was complicated by the fact, that sometimes instead of CRYs one would find photolyases and this had to be taken care of. So instead using BMAL/CYCLE as an indicator would be great and indeed BMAL is a great indicator since no *Drosophila*-like clock with a BMAL TAD was identified. The BMAL TAD or a related sequence also doesn't exist in any other protein, so a BLAST search is all that is needed to identify if a *Drosphila*-like clock is used. This of course is only the case if the gene is correctly annotated/assembled and not wrongly *cycle*-like. Confidence can be increased by including *per* and searching for a CRY binding domain (CBD) which is also only present if a *cry-m* is present. Unfortunately, the *per* sequences are the least reliable in this study because often only fragments are found. Thirdly, the presence of the CLOCK PAS-B HI-loop tryptophan (W362 in the mouse) can be used. The *clock* gene is usually well annotated and assembled but care should be taken to not accidently search for this tryptophan in *bmal*, *arnt* or *per* PAS-B domains. The best is to use all four indicators: CRY-m/CRY-d, BMAL TAD, PER CBD, CLOCK PAS-B W presence as summarized in Table 33.

Table 33: Composition of clock networks

Clock network	CRY-d	CRY-m	BMAL TAD	PER CBD	CLOCK PAS-B W
Drosophila-like	+	-	-	-	-
Honey bee-like	-	+	+	+	+
Monarch-like	+	+	+	+	+

I would argue that looking for interaction site instead of gene presence allows to learn more about the systems. It partially reduces the number of species that can be analyzed but in the genomic era we often anyway don't want to work with insufficiently sequenced species in molecular biology. Instead,

the genome assembly of a close relative can be used. Many genomes are sequenced every year and at ever better quality, it is mostly a question of time to get high quality data of the favorite species or a close relative. Importantly, we are already at the point where we are overwhelmed with the vast amount of publicly available data and need the training and tools to dig deep into this treasure trove.

Interaction site knowledge allows to predict if a certain protein interaction is likely present. The honeybee for example is an incredibly important organism economically but also considered a model organism even tough genetic modification is off limits. The important interaction of CLOCK:MLL1 seems to be lost in the honey bee due to a loss of exon 19 (Figure 24). This was not reported before to my knowledge but is a very rare event in the dataset. Another example is the apparent conservation of the disulfide bridge in CRY-m which was so far only described for mammalian CRY-m but is present in all CRY-m besides a few single species and Chelicerata which are special group themselves. While the functional relevance of the Zn^{2+} interface in the absence of the cognate motif in PER is unclear it is still worth investigating. In the case of linear interaction motifs, the extra effort is worth to gain a deeper understanding of the systems studied (Figure 24).

5.3.3 Conservation of PER:BMAL competition

The PER CBD/BMAL TAD interaction is conserved between monarch and mouse. The PER tail sequence alignment in Figure 25 hints towards a broader conservation because the insect tails are conserved and contain a MI[F/Y]EE[D/N]P[F/L/I]PPP motif. In Chordata conservation is much lower and the motif is [V/M/I/L]EGGG for PER1 which have a distinct glycine rich region after the negative charge. The AlphaFold scores pLDDT and PAE are higher for the insect conserved YEEDPPPP motif which according to the prediction binds to the bottom of the primary pocket as shown in Figure 37. The scores could mean that more co-evolutionary signal was available due to many sequenced insect genomes but would also indicate a higher confidence in the interaction. This would further hint towards an insect wide conservation of the PER:BMAL competition for CRY-m binding. The motif is not present in Chelicerata. The motif further hints towards a tail-based repression as shown for mammalian PER1 which might be conserved based on the similar composition. While the exact mechanism of repression might have changed, the necessity of PER/BMAL competition seems to be conserved. This further hints towards a greater importance of the early repressive phase, i.e. CLOCK hyperphosphorylation and detachment instead of co-activator shedding. This mechanism appears sufficient for repression in *Drosophila* according to the available literature.

5.3.4 Need for multiple CRY-m and PERs

Looking from the human perspective on the circadian clock one might wonder how insects live with just one CRY-m and PER. But arguably the other perspective is more interesting. There are more animals outside the Chordata lineage by number and weight. Their network structure is older and more

widespread. Chordata are the exception. For some reason early in Chordata evolution the genes for CRY-m and PER duplicated and gained new functions. I would argue that CRY2 is the innovation because I expect multistep repression in insects due to PER/BMAL competition for CRY-m. CRY2 doesn't occur alone on BMAL:CLOCK, while CRY1 does as I would also expect for insect CRY-m. CRY2 is proposed to fine-tune the clock by delaying the time at which the more potent repressor CRY1 can bind. In *Danio rerio* two more CRY-m are present but this duplications are lineage-specific.

The Chordata PERs all gained Zn^{2+} binding before duplicating. This interaction is described to be important in correct circadian period and metabolic regulation, as disruption of the motif led to longer periods and hypoinsulinemia[98]. The sequence similarity between PER1 tails shows nicely (Figure 25) that the *per* gene duplication happened before the split into multiple species and in principle hints that results from mouse PER1 are also transferable to *Anolis* or *Danio* PER1. The presence of the third *per* gene in *Danio* and mouse indicates presence of the full set in the last common ancestor with some lineage specific losses as described also by Kotwica-Rolinska and Chodáková (2022)[169]. Other functions such as WIN and WBM sites seem to have been acquired later as they are not present in all chordates and paralogs. These motifs are also relatively small and evolve easier from drift and get fixed than folded domains.

Based on the discussion above on PER/BMAL competition I would argue that insect PERs are probably the closest in function to PER1. Importantly, insect and chordate PER split a long time before chordate *per* gene duplication and this comparison cannot be based on sequence and inheritance but purely on function within the gene regulatory network of the circadian clock. The same is true for this comparison in CRY-m.

5.3.5 Conservation of FASP/PER short

While the FASP region is not conserved in insects as shown in Figure 41 and Table 32, the *Drosophila* per^{s} (per short) and per^{SD} (per short downstream) regions are conserved. Both regions are lying just upstream of CK1 binding sites conserved in insect and mammals. This hints towards convergent evolution for PER control. While PER had to carry CK1 to CLOCK (ancestral function), CK1 binding lead at the root of the mammalian and insect lineages to regulation of PER stability by CK1 phosporylation which was an adaptive advantage leading to fixation. Temperature compensation is largely based on the PER phosphoswitch[40,203] and is an important hallmark of circadian clocks. The usage of different regions suggests convergent evolution.

5.3.6 The curios case of higher Diptera

Interestingly, the *Drosophila*-like clock is very unique and only used by very close relatives of the fruit fly in the Diptera order. On the other hand, CRY-d was lost multiple times independently. *Drosophilas*

uniqueness is clear to many evolutionary biologists in chronobiology but not necessarily clear to everyone. While *Drosophila* sure offers an amazing number of methodological possibilities it also makes it an outsider which is too specialized to easily apply knowledge on other systems. Even tough an incredible amount of research was and continues to be done on *Drosophila* in chronobiology we still lack knowledge on basic questions like how transactivation is achieved, how does TIM-d bind PER and how does PER interact with CLOCK. Hints come from recent bioinformatic studies showing that *Drosophila* has a unique set of casein kinase genes and unique DBT[193]. By comparison with the ancestral monarch-like system is seems that *Drosophila* followed a different route to clock specialization. PER is a CRY-m independent repressor instead of a signal integration module. Stabilization by TIM-d was conserved but is not a new function.

While the CLOCK polyQ motif was discussed to be a transactivation site compensating for TAD loss, such a motif is present in most Choradata and also other Diptera with an ancestral clock network such as *Anopheles*. Importantly it is present in all three *Drosophila*-like clock users in the dataset. The polyQ motif and exon 19 might be the only sites for transactivation but it is still surprising that no deletion mutant was cloned to test this long-standing hypothesis.

5.3.7 The polyQ in Anopheles

Interest sparked by the *Drosophila* CLOCK polyQ motif led to the observation that polyQ motifs are found in more clock proteins and most obviously in almost all *Anopheles* clock proteins such as CLOCK, CRY and PER. The mosquito proteins also are usually the longest which is also true for the especially long tail in CK1[193] . Only marine organisms have longer PER tails. Arguing about the role of these polyQ motifs in clock proteins, activators and repressors alike, would be highly speculative.

The effect of these long and often polyQ motif containing tails was measured for the CRY-m:BMAL TAD affinity which is reduced to 50% by the tail (tail lengths in Figure 27, affinity in Figure 16). The much shorter (Figure 27) mouse CRY1 tail doesn't have a strong effect (data not shown) even tough it binds to the BMAL1 TAD in crosslinking (Figure 14). The long tail in the monarch CRY-m has only mild effects (tail lengths in Figure 27, Figure 18). The important difference could therefore be the composition.

5.3.8 Other networks

In principle other networks should be possible without any CRY-m or CRY-d. The honeybee-like clock entrains CRY-d independent, and the *Drosophila* clock represses CRY-m independent. The ancestral binding partner TIM-d is apparently sufficient to stabilize PER in absence of CRY-m. CRY-m is also not necessary to facilitate the interaction of CLOCK and PER. Theoretically, it should therefore be possible to find a species lacking CRY-d and therefore having the neuronal circuitry for entrainment, that has a TIM-d and lost CRY-m. Such a species would probably be a non-termite Blattodea species (Figure 24) if

PER retained the CRY-m independent repressive activity demonstrated in the ancestral monarch clock[78]. The same would probably not work in Hymenoptera or termites since they lost TIM-d, and PER probably needs at least TIM-d or CRY-m. The absence of such a Blattodea system hints towards a specialization or loss of function of PER over evolution. Alternatively, a spontaneous loss of CRY-m would probably lead to phase advance and reduced amplitude resulting in a competitive disadvantage and selection against it. If retaining the PER repressive activity is advantageous enough to have been kept, then a CRY-m knock down or knock out should yield e. g. ants with advanced phase under free running conditions and probably reduced amplitude. In the mammalian system such mutants are arrhythmic. PER-only repressive functions were lost probably long ago by drift since they didn't yield a competitive advantage in the gene duplicated system which probably resulted in stronger repression. PER repression in CRY-m absence should also not work well in absence of a stabilizing TIM-d. It would be interesting to study which PERs retained PER-only repressive functions and potentially allow to map the responsible CLOCK binding site in PER.

5.4 AlphaFold

The immense improvement that AlphaFold embodies compared to previously prediction methods cannot be overstated. Usage is extremely easy and straight forward. So much that it can be automated as seen by the integration of proteome wide predictions of single proteins into the uniprot database and the EBI hosted AlphaFold DB.

Currently AlphaFold is lacking the ability to include co-factors and ligands even if they are present in homologous structures. If homologous structures are available it apparently keeps their orientations as seen for the Zn^{2+} binding site here, and other examples in the lab. I expect that keeping an orientation that might strain the structure, such as it is the case in some enzymes, would result in suboptimal scoring in absence of the ligand in the protein. I further expect that AlphaFold will be developed to include more user input such as co-factors, ligands, ions and experimental restraints, such as distances between atoms.

The impressive performance of AlphaFold solved the folding problem probably for many monomeric proteins. Yet, the problem of protein complexes is conceptually different. AlphaFold relies a lot on co-evolution in the form of multiple sequence alignment and the intrinsic sampling by evolution shown in the PDB. Such co-evolutionary signals are not as strong between proteins as within proteins. The mechanisms/energy functions governing protein folding also don't necessaryly apply to multimers.

Here user input would be interesting as seen for RoseTTA-Fold which also uses co-evolutionary signatures from MSAs to predict complexes but required a MSA as an input. The case of cryptochromes is a good example of where AlphaFold or any unsupervised MSA building has to fail. The similarity

between the different CRY classes is high enough to place them automatically into the same MSA. But only one subclass (e. g. CRY-m) will interact with a given protein (e. g. BMAL TAD). Unsupervised MSA building will reduce co-evolutionary signal by including non-CRY-m sequences, thereby increase noise.

I tried to predict structures for MLL1:CLOCK exon 19, CLOCK:PER CCID and CLOCK:BMAL:CRY1. They all result in nice monomer predictions with some other monomer far away. The CLOCK:BMAL heterodimer was well predicted as in the crystal structure but AlphaFold failed to build the CRY1:CLOCK PAS-B interaction which it is able to predict with high confidence if only the CLOCK PAS-B domain and CRY1 are given as inputs. It seems some scoring function is satisfied before all predictable interactions are build.

Importantly, AlphaFold enables many scientists to predict protein and protein complex structures without requiring expert knowledge. It thereby massively democratizes state of the art structural prediction capabilities. Careful interpretation of prediction results stays important and the necessity to do so can be easily overlooked once a nice figure is generated. Usage of benchmarked strategies especially in cases of low prediction scores is a necessity just as inspection of those scores. This can allow to focus on the right spot in a protein interaction early on without massive investment in time and consumables to screen large libraries and producing large amounts of trash and perhaps data that is just that, trash. Instead of never ending bench work the scientific method has to be employed. AlphaFold inspires and simplifies building hypothesis based on molecular mechanisms. Such hypothesis have to be compared to available data/knowledge. The derived model finally has to be tested with the appropriate method and controls in the wet lab and withstand rigorous investigation. I believe this just to be next logical step from the development seen in x-ray crystallography: once an expert field, largely occupied by chemists, it is nowadays a method available to much more people without such expert knowledge. This development was enabled by research, decades of experience and development of fantastic computational tools by experts in the field. While most x-ray experiments are routine nowadays and structures can be solved unsupervised, this allows the experts to focus on the tricky, challenging projects. Structural biology already had the problem before people would use structures without looking at B-factors or electron density, misunderstanding error measures and ignoring such important (meta-)information.

The popularity and ease of use of course pose the risk that important and reliable methods fall into oblivion. Loss of methods like molecular dynamics simulations or experimentally enabled predictions like HADDOCK[204] would be detrimental to the scientific community, especially since they are complementary to the stationary or *ab inito* view imposed by AlphaFold.

5.5 Evolution and ecology

5.5.1 Chrono lifestyles and migration

One motivation to study the clock components in multiple organisms was the hypothesis that different traits would be reflected in the central loop. One prediction was that there might be a difference in the circadian clock between nocturnal and diurnal and crepuscular (active at dusk and dawn) animals. Such a difference cannot be found. This is not very surprising since mouse and humans are nocturnal/crepuscular and diurnal, respectively yet the clock components are extremely similar. Further, migratory animals were expected to have perhaps more complex networks to increase robustness. This is not the case as well since many organisms share the clock network with the migratory monarch butterfly but are not migratory themselves. Even on the neurobiological level the central pacemaker is surprisingly small with just four neurons in the monarch[30]. The additional navigational clock residing in the antenna probably yields the necessary robustness, but such adaptations cannot be studied on the sequence level. It appears that the central loop is mostly conserved, and activity patters and navigation/migration are outputs generated later governed by neuro- instead of molecular biology.

5.5.2 Eusociality/subsociality and superorganismic needs

A difference between eusocial and subsocial organisms was also investigated. Eusociality describes the highest level of sociality and includes division of labor into reproducing and non-reproducing groups, brood care of offspring from other individuals and overlapping generations of adults within a colony. The most famous examples are included in the Hymenoptera order which includes bees, ants, and wasps. Why would eusociality influence circadian rhythms? As introduced for the honeybee the eusocial colony, a superorganism, has different needs than the individual. Superorganismic needs have to supersede the individual's circadian clock. Other examples for eusocial insects are termites (order Blattodea). The naked mole-rat (Mammalia) is also described to be eusocial and some shrimps. Interestingly, the Hymenoptera and certain Blattodea are the largest group of species with a honeybee-like clock lacking the photoreceptor CRY-d as shown in Figure 24. These Blattodea species are all termites, hence eusocial. Both, Hymenoptera and termites have also lost TIM-d which is rare. The other large group without CRY-d and TIM-d are the Chordata which also have gene duplications of CRY-m and PER. Only one termite species retained a TIM-d*, Porotermes quadricolis*. The correlation between eusocial insect orders and their clock structure and lack of TIM-d are striking.

A functional link is of course highly speculative, and the loss of CRY-d ad TIM-d precedes the development of eusociality in these animals because many Hymenoptera species are subsocial. A clock network that allows to be superseded by superorganismic needs, hypothetically could be a condition for eusociality instead of eusociality leading to a different clock network. Funnily, actograms of human

parents with newborns look also very arrhythmic just as worker bees taking care of brood (Figure 5). The chronobiological “brood care” phenotype is similar.

5.5.3 Chelicerata

The Chelicerata include spiders, ticks, mites, scorpions, and horseshoe crabs. This group sticks out in the analysis because they have short PERs lacking conserved tail regions as shown in Figure 25. The Chelicerata are also the only group that lacks a CRY-m disulfide pair and one member, the mite *Tetranychus urticae*, seems to have lost the BMAL TAD. The conserved motif in the PER tail could be the reason for BMAL:PER competition for CRY-m and therefore this competition might be absent in these animals. Further, spiders reportedly have rhythms that are longer, about, or shorter than 24 h or are arrhythmic depending on species and individual (personal communication at GRC Chronobiology 2023). These rhythms are stable. Many spiders are predators. The black widow, *Latrodectus mactans*, shows ultradian activity bouts and hunts diurnal and nocturnal prey. A hypothesis is that a lack of circadian behavior allows the black widow to hunt all its prey and not be restricted by its circadian clock[205]. If this phenotype is based on the clock network is unclear. Unfortunately, nothing can be confidently concluded from this directly for the Chelicerata besides spiders. More species would have to be included in this type of study and investigated for circadian rhythms to be able to deduce anything, but it shows that something is going on in this group.

5.5.4 The curios case of marine animals

One key question was if the habitat would have an influence on the clock (components). While the terrestrial animals in this study occupy all kinds of habitats the clock (components) do not reflect this. On the other hand, the few included marine organisms show certain similarities across phyla. Mollusca (oyster, octopus, sepia, sea hare (slug)), Annelida (*Platynereis dumerilii*) have ancestral clocks and importantly the annelid shows circalunar spawning behavior. The marine Chordata (sea lamprey, lancelet) have a honeybee-like clock in the sense that they lack the CRY-m and PER gene duplications found in the other Chordata in the study. The conserved motif in the PER tail is present and rather Chordata-like than insect-like as shown in Figure 25. The PER tails of these species are not only similar in this regard but are also the longest in the study with more than a 100 aa besides the oyster at just over 50 aa. Sepia and Octopus PER tail are about 200 aa long and *Platynereis* and sea hare PER tails are at about 300 aa as shown in Figure 27. Since the function of a long PER tail is unclear the functional significance of these tails is mysterious. The shorter tails of most other organisms in the study suggests an ancestral shorter tail and therefore lengthening of the tail later in these marine organisms. These tails are not completely intrinsically disordered and might harbor interaction sites for other proteins.

The other extreme is covered in the Hemichordata (acorn worm) and Echinodermata (sea stars and sea urchins) which apparently lack PER all together. These are the only animals identified without PER. PER

plays a very central role in the clock and got more into the focus in recent years. The importance of post-translational modifications (PTMs) for clock function and conservation of phosphorylation cycles for activator repression in circadian clocks are better understood. PER recruits the responsible kinase CK1 to circadian promoters. It is therefore even more surprising that PER is lost completely and nothing special is observed in the other clock components. This could of course be due to a lack of sequencing quality instead of lack of sequence. Generally, the quality of PER sequence data was the worst of the clock components. But in principle a clock mechanism without PER is conceivable and would be a different route in clock specialization. In such a model CRY-d could over TIM-d ensure entrainment by controlling the stability of an unknown factor or directly CRY-m. CRY-m can repress by itself as is easily imaginable. While this might not yield the most robust clock it is conceivable that these animals who live in and on the sea floor are not really influenced by day and night as they don't have much influence on their habitat. This would be even more interesting in the context of deep sea creatures where day-night cycles are even less pronounced.

Importantly, longer PER tails or absence of PERs are only observed in marine animals.

5.5.5 Intertidal zone organism in the spotlight

While the environment in the sea can be very stable compared to the environment on land, the environment between the two is even more changing. Daily, monthly, and half-monthly changes overlap due to earth's rotation and the moon's rotation around earth in the intertidal zone. *Clunio marinus* was already introduced. Unfortunately, it was not included in the original evolutionary study but importantly *Clunio* BMAL has two isoforms A and B. Isoform A is an almost typical insect TAD with a D->N substitution, isoleucine instead of leucine in the C-terminus and lacking the insect typical final proline. The second isoform has a very peculiar C-terminus as shown in Figure 33. The habitat of *Clunio* includes different coasts in Europe with different timing of neap tides requiring adaptation. It was shown that timing is genetically determined by crossing of different strains. Further it was shown that calmodulin-dependent kinase II.1 (CaMKII.1) splice variants are strain-specific, and the gene lies in the most strongly associated region with circadian chronotypes. CaMKII.1 increases CLOCK:BMAL activity. *Clunio* is an example on how the evolution of chronotypes and circadian adaptation to the environment can be studied to find genetic determinants. The same species lives in nearly identical habitats but different time restrains as the variable selective pressure and chronotypes are genetically determined. The intertidal zone is home for many more species with special and interesting adaptations. Why the *Clunio* BMAL TAD is so different remains to be studied.

Organisms with more than just circadian rhythms are understudied and the understanding of for example circalunar clocks is still low.

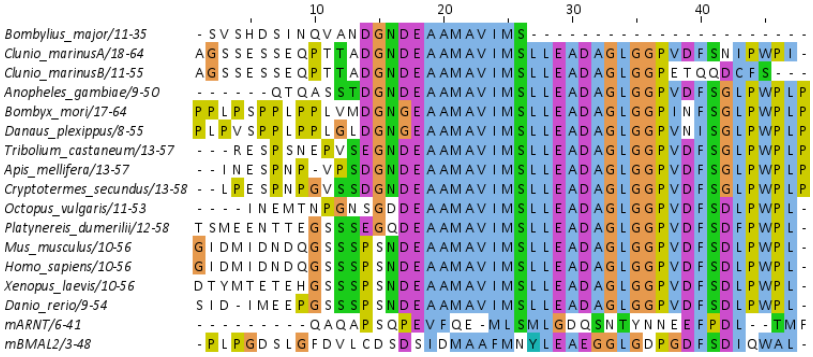

Figure 33: Alignment of Clunio BMAL TAD isoforms with TADs of selected animals. Mammalian BMAL2 and ARNT are shown for comparison. The BMAL TAD C-terminus of Clunio BMAL isoform A is mammal-like because it lacks the C-terminal proline. The insect C-terminus specific D->G substitution is absent in isoform B while isoform A has a unique asparagine (N). D->N is only a charge substitution (negative to neutral) and unlike the substitution to G, does not increase flexibility. The C-terminal leucine (L) to isoleucine (I) exchange in isoform A seems to be important since it is absent in all other species even though it only requires a single nucleotide exchange indicating selection against isoleucine in other species. The C-terminus of the B isoform is a peculiar.

5.6 Conclusion

5.6.1 The CRY-m:BMAL1 TAD interaction

According to the literature the binding site of the activator BMAL TAD on the repressor CRY-m is located at the CRY-m α-helix 22, also called the CC-helix. The literature appears to be biased by the usage of isolated peptides of this helix. Instead, herein the binding site is proposed to be in the primary pocket of the CRY-m PHR. Unfortunately, the data is only suggestive and not conclusive due to the chosen validation approach. Mapping of the binding site by screening of a CRY-m random mutagenesis library by yeast-two-hybrid is recommended as an unbiased, medium throughput approach that can be validated.

Interestingly, not only the interaction itself is conserved in many animal species by sequence homology but even the preference towards a negatively charged BMAL TAD over a neutral one is preserved as demonstrated in CRY-m from different insect orders.

5.6.2 Competitive binding of PERIOD and BMAL1 TAD to repressor CRY

The competition of the co-repressor PER with the activator BMAL for repressor CRY-m binding was demonstrated. The competing epitopes were identified and are different for the mouse paralogs PER1 and PER2. The mechanism is evidently conserved in the monarch butterfly and sequence analysis suggested conservation in all animals. The PER/BMAL competition underlines the importance of

temporal separation and the multi-step process of repression to disassemble the activation complex resulting in stable circadian phase. Notably, earlier reported methodically almost identical biophysical experiments showed a lack of competition for human PER2[96] which is in stark contrast to our findings arising from by multiple approaches including biophysical and cell biological methods. The chordate PER1 paralog is vastly understudied compared to its known essential function within the clock.

5.6.3 Evolutionary analysis of interactions sites in the circadian clock central loop

Apparently, the last common ancestor of animals had a monarch-like clock with all components and complexity. Most animals in the study retained the ancestral structure including Arthropoda, Annelida and Mollusca. CRY-d was lost very early in the Chordata lineage and *cry-m* and *per* genes duplicated in a subgroup of chordates around the time of the Cambrian explosion. Chordata PER gained functions before (Zn^{2+} stabilization) and after (flexible tail for BMAL competition) the gene duplication. Sea urchins, sea stars and acorn worm seem to have lost *per* very on. A diversity of clock networks is present in insects: the ancestral monarch-like, the honeybee-like and the *Drosophila*-like. The honeybee-like network is similar to the mammalian network but importantly comes in two flavors: with TIM-d of unknown function in absence of CRY-d, and without TIM-d which correlates with eusociality. How eusociality depends on the honeybee-like network is unclear. CRY-d and TIM-d were lost independently in two insect orders which later developed eusociality. The networks are based on lineage-specific losses which sometimes happened at the order level (butterflies (Lepidoptera) and wasps (Hymenoptera)). Clock robustness in all networks implies specialization of the retained network components. Convergent evolution is seen in PER proteins: (i) PERs of marine organisms from different phyla have high similarity, and (ii) the phosphorylation-dependent regulation of PER stability evolved apparently independently in chordates and insects.

Medical research frequently studies disease mechanisms (pathophysiology) and learns about physiology because we only learn by comparing different or disturbed systems to understand the significance of the parts. Ecological research learns from studying animals from diverse habitats with diverse phenotypes. Especially the marine environment appears to have an influence on the circadian clock central loop. More challenging habitats such as the intertidal zone which invoke circasemilunar adaptations will allow us to learn more about the circadian clock.

We expect to publish two papers from this thesis, a molecular mechanistic paper on BMAL TAD:CRY-m interaction and competition by PER and a second one on the evolutionary analysis.

6 Bibliography

1. Crosby P, Partch CL. New insights into non-transcriptional regulation of mammalian core clock proteins. *J Cell Sci*. 2020;133(18). doi:10.1242/JCS.241174/226239
2. Rosensweig C, Green CB. Periodicity, repression, and the molecular architecture of the mammalian circadian clock. *Eur J Neurosci*. 2020;51(1):139-165. doi:10.1111/EJN.14254
3. Partch CL. Orchestration of Circadian Timing by Macromolecular Protein Assemblies. Published online 2020. doi:10.1016/j.jmb.2019.12.046
4. Kaiser TS, Poehn B, Szkiba D, et al. The genomic basis of circadian and circalunar timing adaptations in a midge. *Nat 2016 5407631*. 2016;540(7631):69-73. doi:10.1038/nature20151
5. Raible F, Takekata H, Tessmar-Raible K. An overview of monthly rhythms and clocks. *Front Neurol*. 2017;8(MAY):261222. doi:10.3389/fneur.2017.00189
6. Tessmar-Raible K, Raible F, Arboleda E. Another place, another timer: Marine species and the rhythms of life. *BioEssays*. 2011;33(3):165-172. doi:10.1002/BIES.201000096
7. HAUENSCHILD C. Lunar Periodicity. *Cold Spring Harb Symp Quant Biol*. 1960;25:491-497. doi:10.1101/SQB.1960.025.01.051
8. Richmond RH, Hunter CL. Reproduction and recruitment of corals: comparisons among the Caribbean, the tropical Pacific, and the Red Sea. *Mar Ecol Prog Ser*. 1990;60(1):185-203. Accessed August 1, 2023. https://www.int-res.com/articles/meps/60/m060p185.pdf
9. Neumann D. Semilunarperiodische Fortpflanzung von Clunio marinus - Biologische Zeitmessung in der Gezeitenzone - TIB AV-Portal. IWF (Göttingen). doi:10.3203/IWF/C-1091eng
10. ASCHOFF J. Exogenous and Endogenous Components in Circadian Rhythms. *Cold Spring Harb Symp Quant Biol*. 1960;25:11-28. doi:10.1101/SQB.1960.025.01.004
11. Edery I. Circadian rhythms in a nutshell. *Physiol Genomics*. 2000;2000(3):59-74. doi:10.1152/PHYSIOLGENOMICS.2000.3.2.59
12. Bae K, Jin X, Maywood ES, Hastings MH, Reppert SM, Weaver DR. Differential functions of mPer1, mPer2, and mPer3 in the SCN circadian clock. *Neuron*. 2001;30(2):525-536. doi:10.1016/S0896-6273(01)00302-6
13. Ramanathan C, Khan SK, Kathale ND, Xu H, Liu AC. Monitoring cell-autonomous circadian clock rhythms of gene expression using luciferase bioluminescence reporters. *J Vis Exp*. 2012;(67):1-9. doi:10.3791/4234
14. Balsalobre A, Damiola F, Schibler U. A Serum Shock Induces Circadian Gene Expression in Mammalian Tissue Culture Cells. *Cell*. 1998;93(6):929-937. doi:10.1016/S0092-8674(00)81199-X
15. Buhr ED, Yoo SH, Takahashi JS. Temperature as a universal resetting cue for mammalian circadian oscillators. *Science (80-)*. 2010;330(6002):379-385. doi:10.1126/SCIENCE.1195262
16. Rao SM, Mayer AR, Harrington DL. The evolution of brain activation during temporal processing. *Nat Neurosci 2001 43*. 2001;4(3):317-323. doi:10.1038/85191
17. Güler AD, Ecker JL, Lall GS, et al. Melanopsin cells are the principal conduits for rod–cone input to non-image-forming vision. *Nat 2008 4537191*. 2008;453(7191):102-105. doi:10.1038/nature06829

18. Panda S, Provencio I, Tu DC, et al. Melanopsin is required for non-image-forming photic responses in blind mice. *Science (80-)*. 2003;301(5632):525-527. doi:10.1126/SCIENCE.1086179

19. Foster RG, Provencio I, Hudson D, Fiske S, De Grip W, Menaker M. Circadian photoreception in the retinally degenerate mouse (rd/rd). *J Comp Physiol A*. 1991;169(1):39-50. doi:10.1007/BF00198171

20. Berson DM, Dunn FA, Takao M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science (80-)*. 2002;295(5557):1070-1073. doi:10.1126/SCIENCE.1067262

21. Panda S, Sato TK, Castrucci AM, et al. Melanopsin (Opn4) requirement for normal light-induced circadian phase shifting. *Science (80-)*. 2002;298(5601):2213-2216. doi:10.1126/SCIENCE.1076848/

22. Ruby NF, Brennan TJ, Xie X, et al. Role of melanopsin in circadian responses to light. *Science (80-)*. 2002;298(5601):2211-2213. doi:10.1126/SCIENCE.1076701

23. Hattar S, Liao HW, Takao M, Berson DM, Yau KW. Melanopsin-containing retinal ganglion cells: Architecture, projections, and intrinsic photosensitivity. *Science (80-)*. 2002;295(5557):1065-1070. doi:10.1126/SCIENCE.1069609

24. Mano H, Fukada Y. A Median Third Eye: Pineal Gland Retraces Evolution of Vertebrate Photoreceptive Organs†. *Photochem Photobiol*. 2007;83(1):11-18. doi:10.1562/2006-02-24-IR-813

25. Finger AM, Dibner C, Kramer A. Coupled network of the circadian clocks: a driving force of rhythmic physiology. *FEBS Lett*. 2020;594(17):2734-2769. doi:10.1002/1873-3468.13898

26. Hirota T, Fukada Y. Resetting Mechanism of Central and Peripheral Circadian Clocks in Mammals. *https://doi.org/102108/zsj21359*. 2004;21(4):359-368. doi:10.2108/ZSJ.21.359

27. Wehrens SMT, Christou S, Isherwood C, et al. Meal Timing Regulates the Human Circadian System. *Curr Biol*. 2017;27(12):1768-1775.e3. doi:10.1016/J.CUB.2017.04.059

28. Wolff G, Esser KA. Scheduled exercise phase shifts the circadian clock in skeletal Muscle. *Med Sci Sports Exerc*. 2012;44(9):1663-1670. doi:10.1249/MSS.0B013E318255CF4C

29. Youngstedt SD, Elliott JA, Kripke DF. Human circadian phase–response curves for exercise. *J Physiol*. 2019;597(8):2253-2268. doi:10.1113/JP276943

30. Zhu H, Sauman I, Yuan Q, et al. Cryptochromes Define a Novel Circadian Clock Mechanism in Monarch Butterflies That May Underlie Sun Compass Navigation. Mignot E, ed. *PLoS Biol*. 2008;6(1):e4. doi:10.1371/journal.pbio.0060004

31. Haydon MJ, Bell LJ, Webb AAR. Interactions between plant circadian clocks and solute transport. *J Exp Bot*. 2011;62(7):2333-2348. doi:10.1093/JXB/ERR040

32. Heintzen C, Liu Y. The Neurospora crassa Circadian Clock. *Adv Genet*. 2007;58:25-66. doi:10.1016/S0065-2660(06)58002-2

33. Chavan A, Heisler J, Chang YG, Golden SS, Partch CL, LiWang A. Protocols for in vitro reconstitution of the cyanobacterial circadian clock. *Biopolymers*. Published online 2023:e23559. doi:10.1002/BIP.23559

34. Chavan AG, Swan JA, Heisler J, et al. Reconstitution of an intact clock reveals mechanisms of circadian timekeeping. *Science (80-)*. 2021;374(6564). doi:10.1126/SCIENCE.ABD4453

35. Huang N, Chelliah Y, Shan Y, et al. Crystal structure of the heterodimeric CLOCK:BMAL1

transcriptional activator complex. *Science (80-)*. 2012;337(6091):189-194. doi:10.1126/SCIENCE.1222804

36. Michael AK, Stoos L, Crosby P, et al. Cooperation between bHLH transcription factors and histones for DNA access. *Nature*. 2023;619(7969):385-393. doi:10.1038/s41586-023-06282-3

37. Koike N, Yoo SH, Huang HC, et al. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science (80-)*. 2012;338(6105):349-354. doi:10.1126/SCIENCE.1226339

38. Aryal RP, Kwak PB, Tamayo AG, et al. Macromolecular Assemblies of the Mammalian Circadian Clock. *Mol Cell*. 2017;67(5):770-782.e6. doi:10.1016/j.molcel.2017.07.017

39. An Y, Yuan B, Xie P, et al. Decoupling PER phosphorylation, stability and rhythmic expression from circadian clock function by abolishing PER-CK1 interaction. *Nat Commun 2022 131*. 2022;13(1):1-17. doi:10.1038/s41467-022-31715-4

40. Zhou M, Kim JK, Eng GWL, Forger DB, Virshup DM. A Period2 Phosphoswitch Regulates and Temperature Compensates Circadian Period. *Mol Cell*. 2015;60(1):77-88. doi:10.1016/J.MOLCEL.2015.08.022

41. Yoo SH, Mohawk JA, Siepka SM, et al. Competing E3 Ubiquitin Ligases Govern Circadian Periodicity by Degradation of CRY in Nucleus and Cytoplasm. *Cell*. 2013;152(5):1091-1105. doi:10.1016/J.CELL.2013.01.055

42. Xing W, Busino L, Hinds TR, et al. SCFFBXL3 ubiquitin ligase targets cryptochromes at their cofactor pocket. *Nat 2013 4967443*. 2013;496(7443):64-68. doi:10.1038/nature11964

43. Yuan Q, Metterville D, Briscoe AD, Reppert SM. Insect cryptochromes: Gene duplication and loss define diverse ways to construct insect circadian clocks. *Mol Biol Evol*. 2007;24(4):948-955. doi:10.1093/molbev/msm011

44. Peschel N, Chen KF, Szabo G, Stanewsky R. Light-Dependent Interactions between the Drosophila Circadian Clock Factors Cryptochrome, Jetlag, and Timeless. *Curr Biol*. 2009;19(3):241-247. doi:10.1016/J.CUB.2008.12.042

45. Koh K, Zheng X, Sehgal A. JETLAG resets the Drosophila circadian clock by promoting light-induced degradation of TIMELESS. *Science (80-)*. 2006;312(5781):1809-1812. doi:10.1126/SCIENCE.1124951

46. Lin C, Feng S, DeOliveira CC, Crane BR. Cryptochrome–Timeless structure reveals circadian clock timing mechanisms. *Nat 2023 6177959*. 2023;617(7959):194-199. doi:10.1038/s41586-023-06009-4

47. Chiu JC, Ko HW, Edery I. NEMO/NLK Phosphorylates PERIOD to Initiate a Time-Delay Phosphorylation Circuit that Sets Circadian Clock Speed. *Cell*. 2011;145(4):635. doi:10.1016/J.CELL.2011.04.023

48. Chiu JC, Vanselow JT, Kramer A, Edery I. The phospho-occupancy of an atypical SLIMB-binding site on PERIOD that is phosphorylated by DOUBLETIME controls the pace of the clock. *Genes Dev*. 2008;22(13):1758-1772. doi:10.1101/GAD.1682708

49. Garbe DS, Fang Y, Zheng X, et al. Cooperative Interaction between Phosphorylation Sites on PERIOD Maintains Circadian Period in Drosophila. *PLOS Genet*. 2013;9(9):e1003749. doi:10.1371/JOURNAL.PGEN.1003749

50. Beer K, Helfrich-Förster C. Model and Non-model Insects in Chronobiology. *Front Behav Neurosci*. 2020;14(November):1-23. doi:10.3389/fnbeh.2020.601676

51. Lugena AB, Zhang Y, Menet JS, Merlin C. Genome-wide discovery of the daily transcriptome, DNA regulatory elements and transcription factor occupancy in the monarch butterfly brain. *PLoS Genet*. 2019;15(7):e1008265. doi:10.1371/journal.pgen.1008265

52. Takahashi JS. Transcriptional architecture of the mammalian circadian clock. *Nat Rev Genet 2016 183*. 2016;18(3):164-179. doi:10.1038/nrg.2016.150

53. Panda S. Circadian physiology of metabolism. *Science (80-)*. 2016;354(6315):1008-1015. doi:10.1126/SCIENCE.AAH4967

54. Bass J, Lazar MA. Circadian time signatures of fitness and disease. *Science (80-)*. 2016;354(6315):994-999. doi:10.1126/SCIENCE.AAH4965

55. Zhang EE, Liu Y, Dentin R, et al. Cryptochrome mediates circadian regulation of cAMP signaling and hepatic gluconeogenesis. *Nat Med 2010 1610*. 2010;16(10):1152-1156. doi:10.1038/nm.2214

56. Lamia KA, Papp SJ, Yu RT, et al. Cryptochromes mediate rhythmic repression of the glucocorticoid receptor. *Nat 2011 4807378*. 2011;480(7378):552-556. doi:10.1038/nature10700

57. Patke A, Murphy PJ, Onat OE, et al. Mutation of the Human Circadian Clock Gene CRY1 in Familial Delayed Sleep Phase Disorder. *Cell*. 2017;169(2):203-215.e13. doi:10.1016/j.cell.2017.03.027

58. Vanselow K, Vanselow JT, Westermark PO, et al. Differential effects of PER2 phosphorylation: molecular basis for the human familial advanced sleep phase syndrome (FASPS). *Genes Dev*. 2006;20(19):2660-2672. doi:10.1101/GAD.397006

59. Toh KL, Jones CR, He Y, et al. An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science (80-)*. 2001;291(5506):1040-1043. doi:10.1126/SCIENCE.1057499

60. Xu Y, Toh KL, Jones CR, Shin JY, Fu YH, Ptáček LJ. Modeling of a Human Circadian Mutation Yields Insights into Clock Regulation by PER2. *Cell*. 2007;128(1):59-70. doi:10.1016/J.CELL.2006.11.043

61. Narasimamurthy R, Hunt SR, Lu Y, et al. CK1δ/e protein kinase primes the PER2 circadian phosphoswitch. *Proc Natl Acad Sci U S A*. 2018;115(23):5986-5991. doi:10.1073/PNAS.1721076115

62. Parico GCG, Perez I, Fribourgh JL, Hernandez BN, Lee HW, Partch CL. The human CRY1 tail controls circadian timing by regulating its association with CLOCK:BMAL1. *Proc Natl Acad Sci U S A*. 2020;117(45):27971-27979. doi:10.1073/pnas.1920653117

63. Beer K, Bloch G. Circadian plasticity in honey bees. *Biol Clocks*. 2020;42(2):22-26. doi:10.1042/BIO04202002

64. Bloch G, Bar-Shai N, Cytter Y, Green R. Time is honey: circadian clocks of bees and flowers and how their interactions may influence ecological communities. *Philos Trans R Soc B Biol Sci*. 2017;372(1734). doi:10.1098/RSTB.2016.0256

65. Merlin C, Gegear RJ, Reppert SM. Antennal Circadian Clocks Coordinate Sun Compass Orientation in Migratory Monarch Butterflies. *Science (80-)*. 2009;325(5948):1696-1700. doi:10.1126/science.1174562

66. Reppert SM, Guerra PA, Merlin C. Neurobiology of Monarch Butterfly Migration. *Annu Rev Entomol*. 2016;61(1):25-42. doi:10.1146/annurev-ento-010814-020855

67. Sato TK, Yamada RG, Ukai H, et al. Feedback repression is required for mammalian circadian clock function. *Nat Genet 2006 383*. 2006;38(3):312-319. doi:10.1038/ng1745

68. Kiyohara YB, Tagao S, Tamanini F, et al. The BMAL1 C terminus regulates the circadian transcription feedback loop. *Proc Natl Acad Sci U S A*. 2006;103(26):10074-10079. doi:10.1073/PNAS.0601416103

69. Takahata S, Ozaki T, Mimura J, Kikuchi Y, Sogawa K, Fujii-Kuriyama Y. Transactivation mechanisms of mouse clock transcription factors, mClock and mArnt3. *Genes to Cells*. 2000;5(9):739-747. doi:10.1046/J.1365-2443.2000.00363.X

70. Etchegaray JP, Lee C, Wade PA, Reppert SM. Rhythmic histone acetylation underlies transcription in the mammalian circadian clock. *Nat 2003 4216919*. 2002;421(6919):177-182. doi:10.1038/nature01314

71. Xu H, Gustafson CL, Sammons PJ, et al. Cryptochrome 1 regulates the circadian clock through dynamic interactions with the BMAL1 C terminus. *Nat Struct Mol Biol*. 2015;22(6):476-484. doi:10.1038/nsmb.3018

72. Garg A, Orru R, Ye W, et al. Structural and mechanistic insights into the interaction of the circadian transcription factor BMAL1 with the KIX domain of the CREB-binding protein. *J Biol Chem*. 2019;294(45):16604-16619. doi:10.1074/JBC.RA119.009845

73. Gao M, Yang J, Liu S, Su Z, Huang Y. Intrinsically Disordered Transactivation Domains Bind to TAZ1 Domain of CBP via Diverse Mechanisms. *Biophys J*. 2019;117(7):1301-1310. doi:10.1016/J.BPJ.2019.08.026

74. Ogryzko V V., Schiltz RL, Russanova V, Howard BH, Nakatani Y. The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases. *Cell*. 1996;87(5):953-959. doi:10.1016/S0092-8674(00)82001-2

75. Das C, Roy S, Namjoshi S, et al. Binding of the histone chaperone ASF1 to the CBP bromodomain promotes histone acetylation. *Proc Natl Acad Sci U S A*. 2014;111(12):E1072-E1081. doi:10.1073/PNAS.1319122111

76. Katada S, Sassone-Corsi P. The histone methyltransferase MLL1 permits the oscillation of circadian gene expression. *Nat Struct Mol Biol*. 2010;17(12):1414-1421. doi:10.1038/nsmb.1961

77. King DP, Zhao Y, Sangoram AM, et al. Positional Cloning of the Mouse Circadian Clock Gene. *Cell*. 1997;89(4):641-653. doi:10.1016/S0092-8674(00)80245-7

78. Zhang Y, Iiams SE, Menet JS, Hardin PE, Merlin C. TRITHORAX-dependent arginine methylation of HSP68 mediates circadian repression by PERIOD in the monarch butterfly. *Proc Natl Acad Sci U S A*. 2022;119(4):e2115711119. doi:10.1073/PNAS.2115711119

79. Hou Z, Su L, Pei J, Grishin N V., Zhang H. Crystal Structure of the CLOCK Transactivation Domain Exon19 in Complex with a Repressor. *Structure*. 2017;25(8):1187-1194.e3. doi:10.1016/J.STR.2017.05.023

80. Zhao WN, Malinin N, Yang FC, et al. CIPC is a mammalian circadian clock protein without invertebrate homologues. *Nat Cell Biol 2007 93*. 2007;9(3):268-275. doi:10.1038/ncb1539

81. Doi M, Hirayama J, Sassone-Corsi P. Circadian Regulator CLOCK Is a Histone Acetyltransferase. *Cell*. 2006;125(3):497-508. doi:10.1016/J.CELL.2006.03.033

82. Hirayama J, Sahar S, Grimaldi B, et al. CLOCK-mediated acetylation of BMAL1 controls circadian function. *Nat 2007 4507172*. 2007;450(7172):1086-1090. doi:10.1038/nature06394

83. Czarna A, Breitkreuz H, Mahrenholz CC, Arens J, Strauss HM, Wolf E. Quantitative Analyses of Cryptochrome-mBMAL1 Interactions: MECHANISTIC INSIGHTS INTO THE TRANSCRIPTIONAL REGULATION OF THE MAMMALIAN CIRCADIAN CLOCK. *J Biol Chem*. 2011;286(25):22414-22425. doi:10.1074/JBC.M111.244749

84. Zhang Y, Markert MJ, Groves SC, Hardin PE, Merlin C. Vertebrate-like CRYPTOCHROME 2 from monarch regulates circadian transcription via independent repression of CLOCK and BMAL1 activity. *Proc Natl Acad Sci U S A*. 2017;114(36):E7516-E7525. doi:10.1073/pnas.1702014114

85. Sandrelli F, Costa R, Kyriacou CP, Rosato E. Comparative analysis of circadian clock genes in insects. *Insect Mol Biol*. 2008;17(5):447-463. doi:10.1111/j.1365-2583.2008.00832.x

86. Allada R, White NE, So WV, Hall JC, Rosbash M. A Mutant Drosophila Homolog of Mammalian Clock Disrupts Circadian Rhythms and Transcription of period and timeless. *Cell*. 1998;93(5):791-804. doi:10.1016/S0092-8674(00)81440-3

87. Meireles-Filho ACA, Amoretty PR, Souza NA, Kyriacou CP, Peixoto AA. Rhythmic expression of the cycle gene in a hematophagous insect vector. *BMC Mol Biol*. 2006;7(1):1-10. doi:10.1186/1471-2199-7-38

88. Atanesyan L, Günther V, Dichtl B, Georgiev O, Schaffner W. Polyglutamine tracts as modulators of transcriptional activation from yeast to mammals. *Biol Chem*. 2012;393(1-2):63-70. doi:10.1515/BC-2011-252

89. Usdin K, House NCM, Freudenreich CH. Repeat instability during DNA repair: Insights from model systems. *http://dx.doi.org/103109/104092382014999192*. 2015;50(2):142-167. doi:10.3109/10409238.2014.999192

90. Ripperger JA, Schibler U. Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian Dbp transcription and chromatin transitions. *Nat Genet 2006 383*. 2006;38(3):369-374. doi:10.1038/ng1738

91. Rey G, Cesbron F, Rougemont J, Reinke H, Brunner M, Naef F. Genome-Wide and Phase-Specific DNA-Binding Rhythms of BMAL1 Control Circadian Output Functions in Mouse Liver. *PLOS Biol*. 2011;9(2):e1000595. doi:10.1371/JOURNAL.PBIO.1000595

92. Michael AK, Fribourgh JL, Chelliah Y, et al. Formation of a repressive complex in the mammalian circadian clock is mediated by the secondary pocket of CRY1. *Proc Natl Acad Sci U S A*. 2017;114(7):1560-1565. doi:10.1073/PNAS.1615310114

93. Kucera N, Schmalen I, Hennig S, et al. Unwinding the differences of the mammalian PERIOD clock proteins from crystal structure to cellular function. *Proc Natl Acad Sci*. 2012;109(9):3311-3316. doi:10.1073/PNAS.1113280109

94. Hennig S, Strauss HM, Vanselow K, et al. Structural and Functional Analyses of PAS Domain Interactions of the Clock Proteins Drosophila PERIOD and Mouse PERIOD2. Egli M, ed. *PLoS Biol*. 2009;7(4):e1000094. doi:10.1371/journal.pbio.1000094

95. Yildiz Ö, Doi M, Yujnovsky I, et al. Crystal Structure and Interactions of the PAS Repeat Region of the Drosophila Clock Protein PERIOD. *Mol Cell*. 2005;17(1):69-82. doi:10.1016/j.molcel.2004.11.022

96. Fribourgh JL, Srivastava A, Sandate CR, et al. Dynamics at the serine loop underlie differential affinity of cryptochromes for CLOCK:BMAL1 to control circadian timing. *Elife*. 2020;9. doi:10.7554/eLife.55275

97. Chaves I, Yagita K, Barnhoorn S, Okamura H, van der Horst GTJ, 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. doi:10.1128/mcb.26.5.1743-1753.2006

98. Schmalen I, Reischl S, Wallach T, et al. Interaction of Circadian Clock Proteins CRY1 and PER2 Is Modulated by Zinc Binding and Disulfide Bond Formation. *Cell*. 2014;157(5):1203-1215. doi:10.1016/J.CELL.2014.03.057

99. Nangle SN, Rosensweig C, Koike N, et al. Molecular assembly of the period-cryptochrome circadian transcriptional repressor complex. *Elife*. 2014;3(August2014):1-14. doi:10.7554/ELIFE.03674

100. Park HW, Kim ST, Sancar A, Deisenhofer J. Crystal structure of DNA photolyase from Escherichia coli. *Science (80-)*. 1995;268(5219):1866-1872. doi:10.1126/science.7604260

101. Czarna A, Berndt A, Singh HR, et al. Structures of Drosophila Cryptochrome and Mouse Cryptochrome1 Provide Insight into Circadian Function. *Cell*. 2013;153(6):1394-1405. doi:10.1016/J.CELL.2013.05.011

102. Putker M, Crosby P, Feeney KA, et al. Mammalian Circadian Period, but Not Phase and Amplitude, Is Robust Against Redox and Metabolic Perturbations. *Antioxidants Redox Signal*. 2018;28(7):507-520. doi:10.1089/ARS.2016.6911

103. Rosensweig C, Reynolds KA, Gao P, et al. An evolutionary hotspot defines functional differences between CRYPTOCHROMES. *Nat Commun*. 2018;9(1):1-15. doi:10.1038/s41467-018-03503-6

104. Gatfield D, Schibler U. Proteasomes keep the circadian clock ticking. *Science (80-)*. 2007;316(5828):1135-1136. doi:10.1126/SCIENCE.1144165

105. Hirano A, Yumimoto K, Tsunematsu R, et al. FBXL21 Regulates Oscillation of the Circadian Clock through Ubiquitination and Stabilization of Cryptochromes. *Cell*. 2013;152(5):1106-1118. doi:10.1016/J.CELL.2013.01.054

106. Siepka SM, Yoo SH, Park J, et al. Circadian Mutant Overtime Reveals F-box Protein FBXL3 Regulation of Cryptochrome and Period Gene Expression. *Cell*. 2007;129(5):1011-1023. doi:10.1016/J.CELL.2007.04.030

107. Godinho SIH, Maywood ES, Shaw L, et al. The after-hours mutant reveals a role for Fbxl3 in determining mammalian circadian period. *Science (80-)*. 2007;316(5826):897-900. doi:10.1126/SCIENCE.1141138

108. Miller S, Son YL, Aikawa Y, et al. Isoform-selective regulation of mammalian cryptochromes. *Nat Chem Biol*. 2020;16(6):676-685. doi:10.1038/s41589-020-0505-1

109. Miller S, Srivastava A, Nagai Y, Aikawa Y, Tama F, Hirota T. Structural differences in the FAD-binding pockets and lid loops of mammalian CRY1 and CRY2 for isoform-selective regulation. *Proc Natl Acad Sci U S A*. 2021;118(26):e2026191118. doi:10.1073/PNAS.2026191118

110. Miller S, Kesherwani M, Chan P, et al. CRY2 isoform selectivity of a circadian clock modulator with antiglioblastoma efficacy. *Proc Natl Acad Sci U S A*. 2022;119(40):e2203936119. doi:10.1073/PNAS.2203936119

111. Hirota T, Lee JW, St. John PC, et al. Identification of small molecule activators of cryptochrome. *Science (80-)*. 2012;337(6098):1094-1097. doi:10.1126/SCIENCE.1223710

112. Nangle S, Xing W, Zheng N. Crystal structure of mammalian cryptochrome in complex with a small molecule competitor of its ubiquitin ligase. *Cell Res*. 2013;23(12):1417-1419. doi:10.1038/cr.2013.136

113. Anand SN, Maywood ES, Chesham JE, et al. Distinct and Separable Roles for Endogenous CRY1 and CRY2 within the Circadian Molecular Clockwork of the Suprachiasmatic Nucleus, as Revealed by the Fbxl3Afh Mutation. *J Neurosci*. 2013;33(17):7145-7153. doi:10.1523/JNEUROSCI.4950-12.2013

114. Cao X, Yang Y, Selby CP, Liu Z, Sancar A. Molecular mechanism of the repressive phase of the mammalian circadian clock. *Proc Natl Acad Sci U S A*. 2021;118(2). doi:10.1073/PNAS.2021174118

115. Deppisch P, Helfrich-Förster C, Senthilan PR. The Gain and Loss of Cryptochrome/Photolyase Family Members during Evolution. *Genes (Basel)*. 2022;13(9). doi:10.3390/genes13091613

116. Van Der Horst GTJ, Muijtjens M, Kobayashi K, et al. Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nat 1999 3986728*. 1999;398(6728):627-630. doi:10.1038/19323

117. Vitaterna MH, Selby CP, Todo T, et al. Differential regulation of mammalian period genes and circadian rhythmicity by cryptochromes 1 and 2. *Proc Natl Acad Sci U S A*. 1999;96(21):12114-12119. doi:10.1073/pnas.96.21.12114

118. Zylka MJ, Shearman LP, Weaver DR, Reppert SM. Three period Homologs in Mammals: Differential Light Responses in the Suprachiasmatic Circadian Clock and Oscillating Transcripts Outside of Brain. *Neuron*. 1998;20(6):1103-1110. doi:10.1016/S0896-6273(00)80492-4

119. Field MD, Maywood ES, O'Brien JA, Weaver DR, Reppert SM, Hastings MH. Analysis of Clock Proteins in Mouse SCN Demonstrates Phylogenetic Divergence of the Circadian Clockwork and Resetting Mechanisms. *Neuron*. 2000;25(2):437-447. doi:10.1016/S0896-6273(00)80906-X

120. Loop S, Pieler T. Nuclear import of mPER3 in Xenopus oocytes and HeLa cells requires complex formation with mPER1. *FEBS J*. 2005;272(14):3714-3724. doi:10.1111/J.1742-4658.2005.04798.X

121. Chang DC, Reppert SM. A Novel C-Terminal Domain of Drosophila PERIOD Inhibits dCLOCK:CYCLE-Mediated Transcription. *Curr Biol*. 2003;13(9):758-762. doi:10.1016/S0960-9822(03)00286-0

122. Menet JS, Abruzzi KC, Desrochers J, Rodriguez J, Rosbash M. Dynamic PER repression mechanisms in the Drosophila circadian clock: from on-DNA to off-DNA. *Genes Dev*. 2010;24(4):358-367. doi:10.1101/GAD.1883910

123. Nawathean P, Stoleru D, Rosbash M. A Small Conserved Domain of Drosophila PERIOD Is Important for Circadian Phosphorylation, Nuclear Localization, and Transcriptional Repressor Activity. *Mol Cell Biol*. 2007;27(13):5002. doi:10.1128/MCB.02338-06

124. Gekakis N, Saez L, Delahaye-Brown AM, et al. Isolation of timeless by PER Protein Interaction: Defective Interaction Between timeless Protein and Long-Period Mutant PERL. *Science (80-)*. 1995;270(5237):811-815. doi:10.1126/SCIENCE.270.5237.811

125. Saez L, Young MW. Regulation of Nuclear Entry of the Drosophila Clock Proteins Period and Timeless. *Neuron*. 1996;17(5):911-920. doi:10.1016/S0896-6273(00)80222-6

126. Ünsal-Kaçmaz K, Mullen TE, Kaufmann WK, Sancar A. Coupling of Human Circadian and Cell Cycles by the Timeless Protein. *Mol Cell Biol*. 2005;25(8):3109-3116. doi:10.1128/mcb.25.8.3109-3116.2005

127. Kurien P, Hsu PK, Leon J, et al. TIMELESS mutation alters phase responsiveness and causes advanced sleep phase. *Proc Natl Acad Sci U S A*. 2019;116(24):12045-12053. doi:10.1073/pnas.1819110116

128. Witosch J, Wolf E, Mizuno N. Architecture and ssDNA interaction of the Timeless-Tipin-RPA complex. *Nucleic Acids Res*. 2014;42(20):12912-12927. doi:10.1093/NAR/GKU960

129. Benna C, Bonaccorsi S, Wülbeck C, et al. Drosophila timeless2 Is Required for Chromosome Stability and Circadian Photoreception. *Curr Biol*. 2010;20(4):346-352. doi:10.1016/j.cub.2009.12.048

130. Kotwica-Rolinska J, Damulewicz M, Chodakova L, Kristofova L, Dolezel D. Pigment Dispersing Factor Is a Circadian Clock Output and Regulates Photoperiodic Response in the Linden Bug, Pyrrhocoris apterus. *Front Physiol*. 2022;13(April):1-18. doi:10.3389/fphys.2022.884909

131. Philpott JM, Narasimamurthy R, Ricci CG, et al. Casein kinase 1 dynamics underlie substrate selectivity and the PER2 circadian phosphoswitch. *Elife*. 2020;9. doi:10.7554/ELIFE.52343

132. Shinohara Y, Koyama YM, Ukai-Tadenuma M, et al. Temperature-Sensitive Substrate and Product Binding Underlie Temperature-Compensated Phosphorylation in the Clock. *Mol Cell*. 2017;67(5):783-798.e20. doi:10.1016/J.MOLCEL.2017.08.009

133. He Q, Cha J, He Q, Lee HC, Yang Y, Liu Y. CKI and CKII mediate the FREQUENCY-dependent phosphorylation of the WHITE COLLAR complex to close the Neurospora circadian negative feedback loop. *Genes Dev*. 2006;20(18):2552-2565. doi:10.1101/GAD.1463506

134. He Q, Liu Y. Molecular mechanism of light responses in Neurospora: from light-induced transcription to photoadaptation. *Genes Dev*. 2005;19(23):2888-2899. doi:10.1101/GAD.1369605

135. Schafmeier T, Haase A, Káldi K, Scholz J, Fuchs M, Brunner M. Transcriptional Feedback of Neurospora Circadian Clock Gene by Phosphorylation-Dependent Inactivation of Its Transcription Factor. *Cell*. 2005;122(2):235-246. doi:10.1016/J.CELL.2005.05.032

136. Wang B, Kettenbach AN, Zhou X, Loros JJ, Dunlap JC. The Phospho-Code Determining Circadian Feedback Loop Closure and Output in Neurospora. *Mol Cell*. 2019;74(4):771-784.e3. doi:10.1016/J.MOLCEL.2019.03.003

137. Lee C, Weaver DR, Reppert SM. Direct Association between Mouse PERIOD and CKIε Is Critical for a Functioning Circadian Clock. *Mol Cell Biol*. 2004;24(2):584-594. doi:10.1128/MCB.24.2.584-594.2004

138. Eide EJ, Woolf MF, Kang H, et al. Control of Mammalian Circadian Rhythm by CKIε-Regulated Proteasome-Mediated PER2 Degradation. *Mol Cell Biol*. 2005;25(7):2795-2807. doi:10.1128/mcb.25.7.2795-2807.2005

139. Philpott JM, Freeberg AM, Park J, et al. PERIOD phosphorylation leads to feedback inhibition of CK1 activity to control circadian period. *Mol Cell*. 2023;83(10):1677-1692.e8. doi:10.1016/J.MOLCEL.2023.04.019

140. Kivimäe S, Saez L, Young MW. Activating PER Repressor through a DBT-Directed Phosphorylation Switch. *PLOS Biol*. 2008;6(7):e183. doi:10.1371/JOURNAL.PBIO.0060183

141. Young Kim E, Wan Ko H, Yu W, Hardin PE, Edery I. A DOUBLETIME Kinase Binding Domain on the Drosophila PERIOD Protein Is Essential for Its Hyperphosphorylation, Transcriptional Repression, and Circadian Clock Function. *Mol Cell Biol*. 2007;27(13):5014-5028. doi:10.1128/MCB.02339-06

142. Chen R, Schirmer A, Lee Y, et al. Rhythmic PER Abundance Defines a Critical Nodal Point for Negative Feedback within the Circadian Clock Mechanism. *Mol Cell*. 2009;36(3):417-430. doi:10.1016/J.MOLCEL.2009.10.012

143. Reischl S, Vanselow K, Westermark PO, et al. β-TrCP1-Mediated Degradation of PERIOD2 Is Essential for Circadian Dynamics. *http://dx.doi.org/101177/0748730407303926*. 2007;22(5):375-386. doi:10.1177/0748730407303926

144. Ohsaki K, Oishi K, Kozono Y, Nakayama K, Nakayama KI, Ishida N. The Role of β-TrCP1 and β-TrCP2 in Circadian Rhythm Generation by Mediating Degradation of Clock Protein PER2. *J Biochem*. 2008;144(5):609-618. doi:10.1093/JB/MVN112

145. Yoshitane H, Takao T, Satomi Y, Du N-H, Okano T, Fukada Y. Roles of CLOCK Phosphorylation in Suppression of E-Box-Dependent Transcription. *https://doi.org/101128/MCB01864-08*. 2023;29(13):3675-3686. doi:10.1128/MCB.01864-08

146. Spengler ML, Kuropatwinski KK, Schumer M, Antoch MP. A serine cluster mediates BMAL1-dependent CLOCK phosphorylation and degradation. *http://dx.doi.org/104161/cc82410273*. 2009;8(24):4138-4146. doi:10.4161/CC.8.24.10273

147. Emery P, Stanewsky R, Helfrich-Förster C, Emery-Le M, Hall JC, Rosbash M. Drosophila CRY is a deep brain circadian photoreceptor. *Neuron*. 2000;26(2):493-504. doi:10.1016/S0896-6273(00)81181-2

148. Stanewsky R, Kaneko M, Emery P, et al. The cry(b) mutation identifies cryptochrome as a circadian photoreceptor in Drosophila. *Cell*. 1998;95(5):681-692. doi:10.1016/S0092-8674(00)81638-4

149. Griffin EA, Staknis D, Weitz CJ. Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. *Science (80-)*. 1999;286(5440):768-771. doi:10.1126/SCIENCE.286.5440.768

150. Busza A, Emery-Le M, Rosbash M, Emery P. Roles of the two Drosophila CRYPTOCHROME structural domains in circadian photoreception. *Science (80-)*. 2004;304(5676):1503-1506. doi:10.1126/SCIENCE.1096973

151. Ceriani MF, Darlington TK, Staknis D, et al. Light-dependent sequestration of TIMELESS by CRYPTOCHROME. *Science (80-)*. 1999;285(5427):553-556. doi:10.1126/SCIENCE.285.5427.553

152. Dissel S, Codd V, Fedic R, et al. A constitutively active cryptochrome in Drosophila melanogaster. *Nat Neurosci 2004 78*. 2004;7(8):834-840. doi:10.1038/nn1285

153. Lin F-J, Song W, Meyer-Bernstein E, Naidoo N, Sehgal A. Photic Signaling by Cryptochrome in the Drosophila Circadian System. *https://doi.org/101128/MCB21217287-72942001*. 2023;21(21):7287-7294. doi:10.1128/MCB.21.21.7287-7294.2001

154. Gegear RJ, Casselman A, Waddell S, Reppert SM. Cryptochrome mediates light-dependent magnetosensitivity in Drosophila. *Nature*. 2008;454(7207):1014-1018. doi:10.1038/nature07183

155. Stoneham AM, Gauger EM, Porfyrakis K, Benjamin SC, Lovett BW. A New Type of Radical-Pair-Based Model for Magnetoreception. *Biophys J*. 2012;102(5):961-968. doi:10.1016/J.BPJ.2012.01.007

156. Guerra PA, Gegear RJ, Reppert SM. A magnetic compass aids monarch butterfly migration. *Nat Commun*. 2014;5(1):4164. doi:10.1038/ncomms5164

157. Zoltowski BD, Chelliah Y, Wickramaratne A, et al. Chemical and structural analysis of a photoactive vertebrate cryptochrome from pigeon. *Proc Natl Acad Sci U S A*. 2019;116(39):19449-19457. doi:10.1073/PNAS.1907875116

158. Franz S, Ignatz E, Wenzel S, et al. Structure of the bifunctional cryptochrome aCRY from Chlamydomonas reinhardtii. *Nucleic Acids Res*. 2018;46(15):8010-8022.

doi:10.1093/NAR/GKY621

159. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189-1191. doi:10.1093/BIOINFORMATICS/BTP033

160. Pettersen EF, Goddard TD, Huang CC, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*. 2021;30(1):70-82. doi:10.1002/PRO.3943

161. Goddard TD, Huang CC, Meng EC, et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci*. 2018;27(1):14-25. doi:10.1002/PRO.3235

162. Madeira F, Pearce M, Tivey ARN, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res*. 2022;50(W1):W276-W279. doi:10.1093/NAR/GKAC240

163. Kumar S, Suleski M, Craig JM, et al. TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol Biol Evol*. 2022;39(8). doi:10.1093/MOLBEV/MSAC174

164. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49(W1):W293-W296. doi:10.1093/NAR/GKAB301

165. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42(W1):W320-W324. doi:10.1093/NAR/GKU316

166. Elisabeth Gasteiger, Christine Hoogland, Alexandre Gattiker, Séverine Duvaud, Marc R. Wilkins, Ron D. Appel and AB. *The Proteomics Protocols Handbook*. Humana Press; 2005. doi:10.1385/1592598900

167. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*. 1990;18(20):6097-6100. doi:10.1093/NAR/18.20.6097

168. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A Sequence Logo Generator. *Genome Res*. 2004;14(6):1188-1190. doi:10.1101/GR.849004

169. Kotwica-Rolinska J, Chodáková L, Smýkal V, et al. Loss of Timeless Underlies an Evolutionary Transition within the Circadian Clock. *Mol Biol Evol*. 2022;39(1). doi:10.1093/MOLBEV/MSAB346

170. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nat 2021 5967873*. 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2

171. Evans R, O'Neill M, Pritzel A, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*. Published online March 10, 2022:2021.10.04.463034. doi:10.1101/2021.10.04.463034

172. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods 2009 65*. 2009;6(5):343-345. doi:10.1038/nmeth.1318

173. Gibson DG, Glass JI, Lartigue C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science (80-)*. 2010;329(5987):52-56. doi:10.1126/SCIENCE.1190719

174. Scholz J, Besir H, Strasser C, Suppmann S. A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. *BMC Biotechnol*. 2013;13(1):12. doi:10.1186/1472-6750-13-12

175. Simon MA, Ecsédi P, Kovács GM, et al. High-throughput competitive fluorescence polarization assay reveals functional redundancy in the S100 protein family. *FEBS J*. 2020;287(13):2834-2846. doi:10.1111/FEBS.15175

176. Roehrl MHA, Wang JY, Wagner G. A general framework for development and data analysis of competitive high-throughput screens for small-molecule inhibitors of protein-protein interactions by fluorescence polarization. *Biochemistry*. 2004;43(51):16056-16066. doi:10.1021/BI048233G

177. Conrady MC, Suarez I, Gogl G, et al. Structure of High-Risk Papillomavirus 31 E6 Oncogenic Protein and Characterization of E6/E6AP/p53 Complex Formation. *J Virol*. 2020;95(2). doi:10.1128/JVI.00730-20

178. Tomohiro T, Hashimoto M, Hatanaka Y. Cross-linking chemistry and biology: development of multifunctional photoaffinity probes. *Chem Rec*. 2005;5(6):385-395. doi:10.1002/TCR.20058

179. Hashimoto M, Hatanaka Y. Recent Progress in Diazirine-Based Photoaffinity Labeling. *European J Org Chem*. 2008;2008(15):2513-2523. doi:10.1002/EJOC.200701069

180. Rappsilber J, Ishihama Y, Mann M. Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem*. 2003;75(3):663-670. doi:10.1021/AC026117I

181. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol 2008 2612*. 2008;26(12):1367-1372. doi:10.1038/nbt.1511

182. Lenz S, Giese SH, Fischer L, Rappsilber J. In-Search Assignment of Monoisotopic Peaks Improves the Identification of Cross-Linked Peptides. *J Proteome Res*. 2018;17(11):3923-3931. doi:10.1021/ACS.JPROTEOME.8B00600

183. Mendes ML, Fischer L, Chen ZA, et al. An integrated workflow for crosslinking mass spectrometry. *Mol Syst Biol*. 2019;15(9):e8994. doi:10.15252/MSB.20198994

184. Lee C, Etchegaray JP, Cagampang FRA, Loudon ASI, Reppert SM. Posttranslational mechanisms regulate the mammalian circadian clock. *Cell*. 2001;107(7):855-867. doi:10.1016/S0092-8674(01)00610-9

185. Brown SA, Ripperger J, Kadener S, et al. PERIOD1-associated proteins modulate the negative limb of the mammalian circadian oscillator. *Science (80-)*. 2005;308(5722):693-696. doi:10.1126/SCIENCE.1107373

186. Duong HA, Robles MS, Knutti D, Weitz CJ. A molecular mechanism for circadian clock negative feedback. *Science (80-)*. 2011;332(6036):1436-1439. doi:10.1126/SCIENCE.1196766

187. 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. doi:10.1074/JBC.M111.254680

188. Shearman LP, Sriram S, Weaver DR, et al. Interacting molecular loops in the mammalian circadian clock. *Science (80-)*. 2000;288(5468):1013-1019. doi:10.1126/SCIENCE.288.5468.1013

189. Ye R, Selby CP, Chiou YY, Ozkan-Dagliyan I, Gaddameedhi S, Sancar A. Dual modes of CLOCK:BMAL1 inhibition mediated by Cryptochrome and Period proteins in the mammalian circadian clock. *Genes Dev*. 2014;28(18):1989-1998. doi:10.1101/GAD.249417.114

190. Kume K, Zylka MJ, Sriram S, et al. mCRY1 and mCRY2 Are Essential Components of the Negative Limb of the Circadian Clock Feedback Loop. *Cell*. 1999;98(2):193-205. doi:10.1016/S0092-8674(00)81014-4

191. Lee C. unpublished AFmulti benchmarking. Published online 2023.

192. Jones CR, Huang AL, Ptáček LJ, Fu YH. Genetic basis of human circadian rhythm disorders. *Exp Neurol*. 2013;243:28-33. doi:10.1016/J.EXPNEUROL.2012.07.012

193. Thakkar N, Giesecke A, Bazalova O, et al. Evolution of casein kinase 1 and functional analysis of new doubletime mutants in Drosophila. *Front Physiol*. 2022;13(December):1-17. doi:10.3389/fphys.2022.1062632

194. Hosoda H, Kato K, Asano H, et al. CBP/p300 is a cell type-specific modulator of CLOCK/BMAL1-mediated transcription. *Mol Brain*. 2009;2(1):1-18. doi:10.1186/1756-6606-2-34/FIGURES/10

195. Lee Y, Lee J, Kwon I, et al. Coactivation of the CLOCK–BMAL1 complex by CBP mediates resetting of the circadian clock. *J Cell Sci*. 2010;123(20):3547-3557. doi:10.1242/JCS.070300

196. Mei Y, Jing D, Tang S, et al. InsectBase 2.0: a comprehensive gene resource for insects. *Nucleic Acids Res*. 2022;50(D1):D1040-D1045. doi:10.1093/NAR/GKAB1090

197. Gustafson CL, Parsley NC, Asimgil H, et al. A Slow Conformational Switch in the BMAL1 Transactivation Domain Modulates Circadian Rhythms. *Mol Cell*. 2017;66(4):447-457.e7. doi:10.1016/J.MOLCEL.2017.04.011

198. Shanware NP, Hutchinson JA, Kim SH, Zhan L, Bowler MJ, Tibbetts RS. Casein Kinase 1-dependent Phosphorylation of Familial Advanced Sleep Phase Syndrome-associated Residues Controls PERIOD 2 Stability. *J Biol Chem*. 2011;286(14):12766-12774. doi:10.1074/JBC.M111.224014

199. Börgel A, Rawleigh A, Conrady M, et al. A structural competition involving WDR5 times circadian oscillations. *bioRxiv*. Published online June 5, 2023:2023.06.05.543739. doi:10.1101/2023.06.05.543739

200. Peres R, Amaral FG, Marques AC, Neto JC. Melatonin Production in the Sea Star Echinaster brasiliensis (Echinodermata). *https://doi.org/101086/BBLv226n2p146*. 2014;226(2):146-151. doi:10.1086/BBLV226N2P146

201. Hemsley MJ, Mazzotta GM, Mason M, et al. Linear motifs in the C-terminus of D. melanogaster cryptochrome. *Biochem Biophys Res Commun*. 2007;355(2):531-537. doi:10.1016/j.bbrc.2007.01.189

202. Akashi M, Okamoto A, Tsuchiya Y, Todo T, Nishida E, Node K. A Positive Role for PERIOD in Mammalian Circadian Gene Expression. *Cell Rep*. 2014;7(4):1056-1064. doi:10.1016/j.celrep.2014.03.072

203. Hu Y, Liu X, Lu Q, et al. Frq-ck1 interaction underlies temperature compensation of the neurospora circadian clock. *MBio*. 2021;12(3). doi:10.1128/MBIO.01425-21

204. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*. 2003;125(7):1731-1737. doi:10.1021/JA026939X

205. Mah A, Ayoub N, Toporikova N, Jones TC, Moore D. Locomotor activity patterns in three spider species suggest relaxed selection on endogenous circadian period and novel features of chronotype. *J Comp Physiol A Neuroethol Sensory, Neural, Behav Physiol*. 2020;206(4):499-515. doi:10.1007/s00359-020-01412-y

Annex

```
# start this script from an empty pymol session
import os

def calculate_structural_metric(self, path):
        """Parse the atoms that are in contact in the predicted model. Contacts are
limited to the distance of 5A between two atoms of residues from different chain.
Color the predicted model by chain and display the residues in contact as sticks

        Returns:
            self.atom_atom_contacts (list of list): A list of atom-atom contacts
stored as list [chain_idA, residueA, residue_indexA, atom_nameA, chain_idB, resi-
dueB, residue_indexB, atom_nameB, distance]
            self.num_res_res_contact (int): Number of residue-residue contacts
        """
        # make a selection of atoms in both chain that have at least one atom with
less than or equal to 5A from any atom from the other chain only used for display?
        selection_line = f"(byres {self} and chain A within 5A of {self} and chain
B) or (byres ranked_0 and chain B within 5A of {self} and chain A)"
        cmd.select(selection=selection_line, name="residues_less_5A")
        # iterate through atom indices in one chain that are less than 5A from any
atom of the other chain and store them in a list
        chainA_contact_indices = []
        #is it possible to include something here to expand to full residues?
        cmd.iterate(f"{self} and chain A within 5A of {self} and chain
B","chainA_contact_indices.append([oneletter,resi,name,index])",space={'chainA_con-
tact_indices':chainA_contact_indices})
        # iterate through the list of indices (atoms of A) and find atoms from the
other chain that is less than 5A away and calculate distance between them
        unique_resi_pair = set()
        for oneletterA,resiA,nameA,indexA in chainA_contact_indices:
            chainB_contact_indices = []
            #make a list of indices from B
            cmd.iterate(f"{self} and chain B within 5A from {self} and index {in-
dexA}","chainB_contact_indices.append([oneletter,resi,name,in-
dex])",space={'chainB_contact_indices':chainB_contact_indices})
            # iterate through the indices from the other chain to calculate their
distance with the index from the previous chain
            for oneletterB,resiB,nameB,indexB in chainB_contact_indices:
                #measure the actual distance of list element of B indices to the
current list A index
                dist = cmd.distance("interface_contacts",f"{self}`{in-
dexA}",f"{self}`{indexB}")
                unique_resi_pair.add((resiA,resiB)) # actually only need this info

        # save unique resi pairs
        with open('unique_pairs.txt', 'a') as f0:
            for pair in unique_resi_pair:
                f0.write(f'{pair}\n')
        print("Unique pairs saved.")

        cmd.reinitialize()
```

Script 1: PyMOL script to identify pairs of interacting residues.

```
def calculate(path):
    ### iterating over the predicitions and moving around the OS ###
    os.chdir(path)
    #path = 'Y:/alphafold/KIX_TAD'
    cwd = os.getcwd()
    fragment_types = ['M', 'N', 'C']
    for fragment_class in fragment_types:
        for i in range(1,11):
            os.chdir(path)
            cwd = os.getcwd()
            path1 = os.path.join(cwd, fragment_class+str(i),'results')
            print('This is path1: ',path1)
            os.chdir(path1)
            for file in os.listdir(os.getcwd()):
                cwd = os.getcwd()
                d = os.path.join(cwd, file, "mCRY1PHRCCE")
                os.chdir(d)
                cwd = os.getcwd()
                print('And after building the filepath: ',cwd)
                cmd.load("ranked_0.pdb")
                #re-enumerate chain B
                if fragment_class == 'C':
                    add = 10*i + 522
                elif fragment_class == 'M':
                    add = 532 - 10*i
                elif fragment_class == 'N':
                    add = 432
                cmd.alter('chain B', f'resi=str(int(resi)+{add})')
                calculate_structural_metric("ranked_0", path)
```

Script 2: Second part of the pair finding script to move through the predictions.

```
import os
import pandas as pd
import numpy as np

### make an empty dataframe ###
data = 0
CRY1 = np.arange(1,506,1)
BMAL = np.arange(433,633,1)
df0 = pd.DataFrame(data, index = BMAL, columns = CRY1 )

### moving around the OS to the data ###
os.chdir('Folder with data')
#load actual data
all_pairs = []
with open("unique_pairs.txt", "r") as file:
    for line in file:
        line = line.rstrip()
        line = eval(line)
        all_pairs.append(line)

# make a dataframe from pair data
df1 = pd.DataFrame.from_records(all_pairs, columns = ['chainA', 'chainB'])
df1 = df1.apply(pd.to_numeric)

# find number of pairs
df2 = df1.groupby(df1.columns.tolist(),as_index=False).size()
df2 = df2.apply(pd.to_numeric)

###calculate the frac of occurence, actually kind of unnecessary...###
df2['size'] = df2['size']/11 #each resi pair can only be present 11 times

###pivot the data for plotting as heatmap###
df2 = df2.pivot(index = 'chainB', columns = 'chainA', values='size')
df2 = df2.fillna(0)

# adding data and empty df to fill in all the non-interacting pairs
result = df0+df2
result = result.fillna(0)

ylabel1 ='mBMAL1'
yticks_interval = tick_interval(595, 633, 10)
sns.heatmap(
    result,
    cmap = sns.cm.rocket_r,
    xticklabels = 20,
     yticklabels = yticks_interval,

    cbar_kws={'label': 'fraction of common contacts with distance <5$\AA$'},
    #annot_kws={"size": 5},
    vmin=0,
    vmax=1,
    ax = ax1
    )
ax1.set_yticklabels(ax1.get_yticklabels(), rotation = 0, fontsize = 10)
ax1.set_title('AlphaFold prediction')
ax1.set_ylabel('mBMAL1 TAD', fontsize = 10)
```

Script 3: Calculation of FCC and plotting of AlphaFold result.

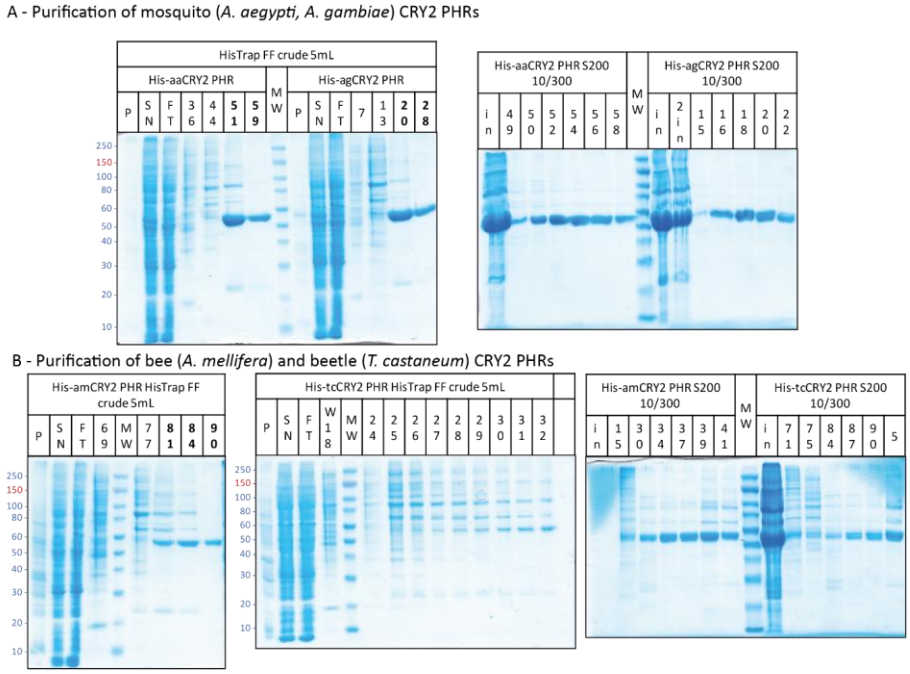

Figure 34: Purification of insect CRY-m PHR.

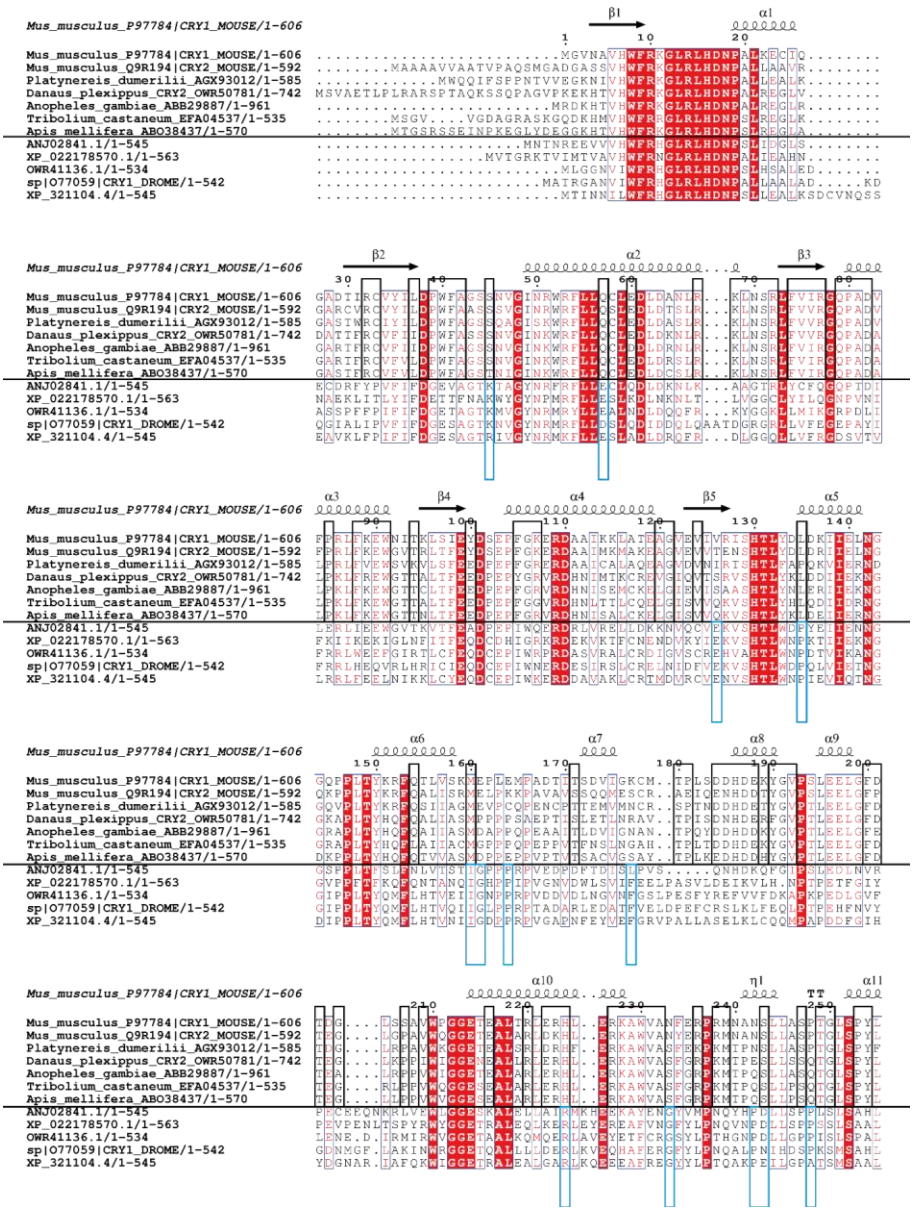

Figure 35: MSA CRY-m, CRY-d. Page 1. The black line indicates the separation between CRY-m (top) and CRY-d (bottom). Black and cyan boxes highlight CRY-m and CRY-d specific conserved positions, respectively. One outlier allowed; chemically similar amino acids are considered conserved as well. The secondary structure annotation is based on an apo mouse CRY1 structure, PDB 6KX4.

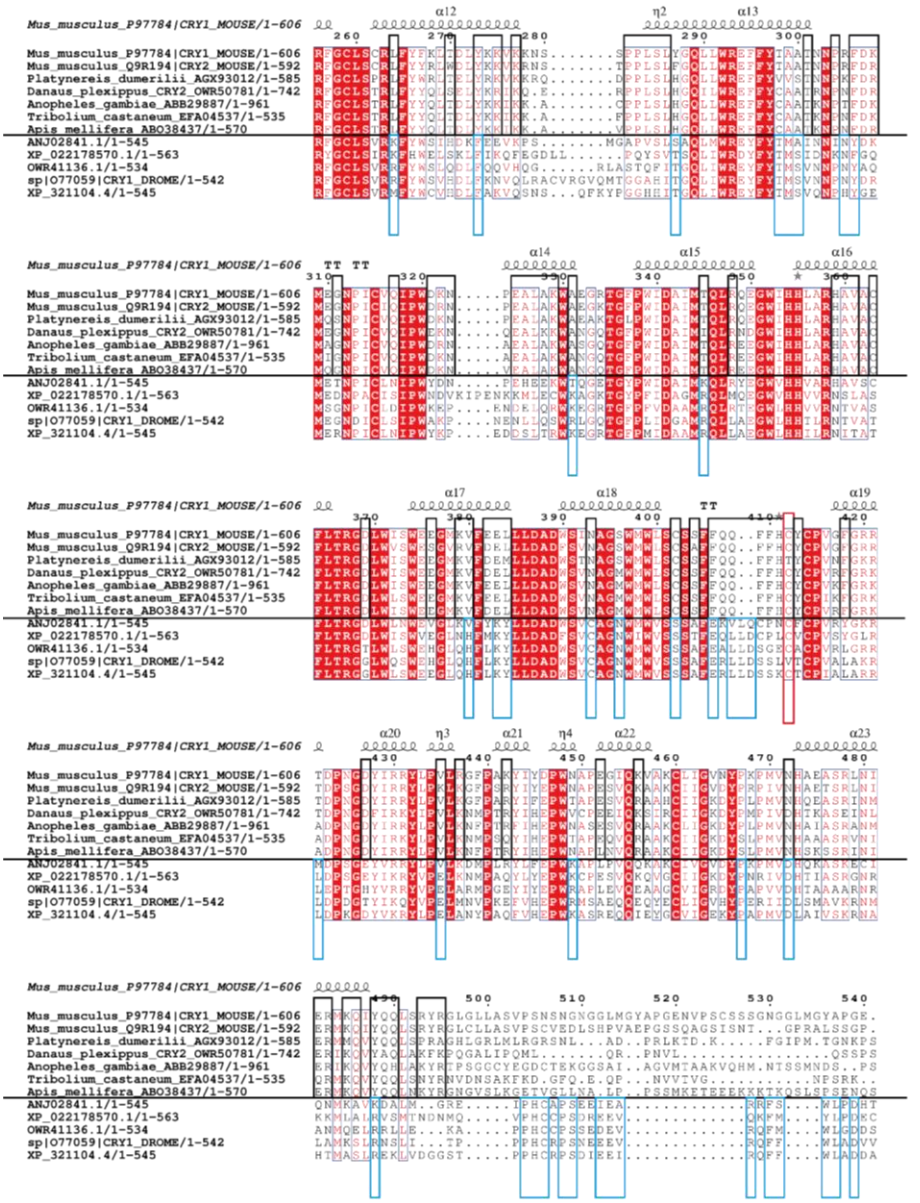

Figure 36: MSA CRY-m, CRY-d. Page2. The black line indicates the separation between CRY-m (top) and CRY-d (bottom). Black and cyan boxes highlight CRY-m and CRY-d specific conserved positions, respectively. One outlier allowed; chemically similar amino acids are considered conserved as well. The secondary structure annotation is based on an apo mouse CRY1 structure, PDB 6KX4. The remaining tails are not shown because of the low conservation this would waste paper without information.

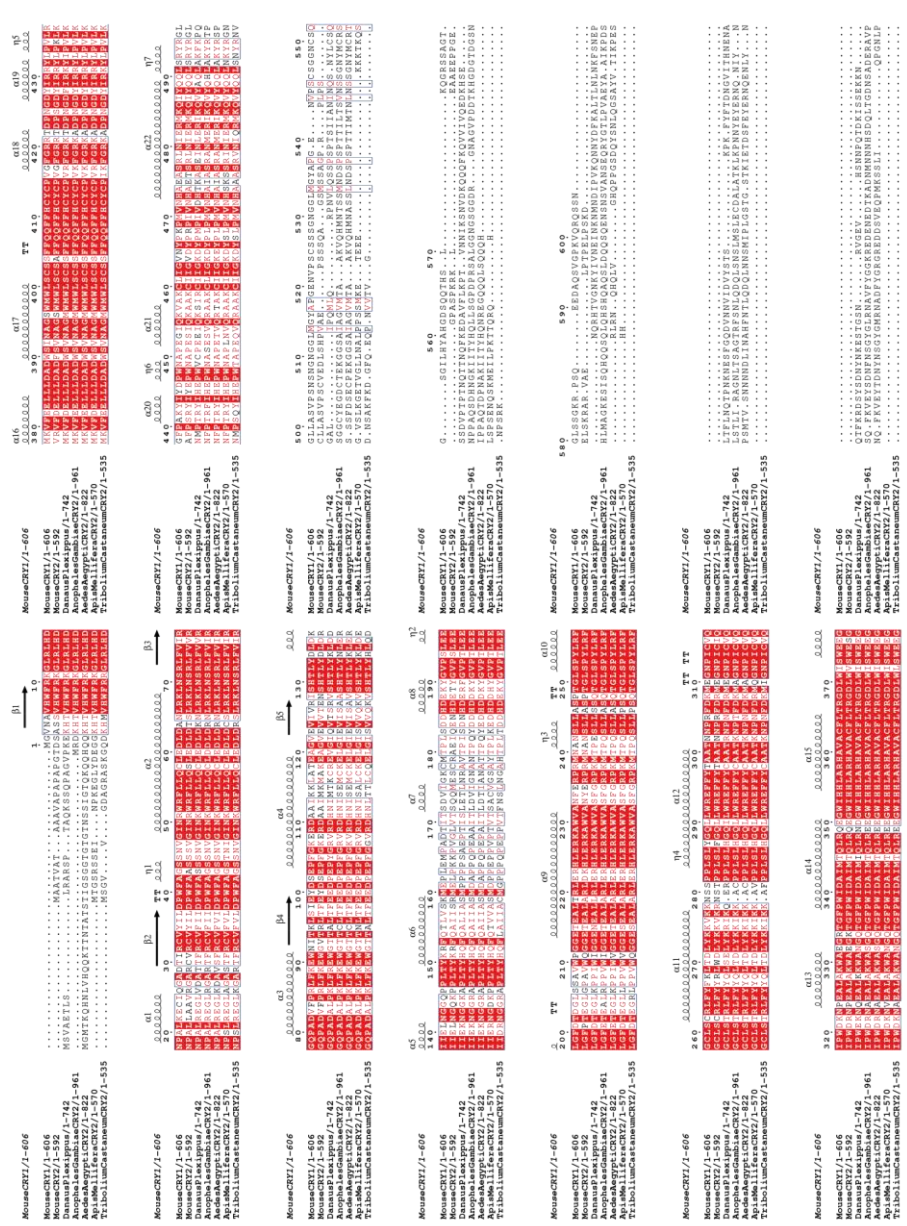

Figure 37: Multiple sequence alignment of mouse and insect CRY-m used in this thesis.

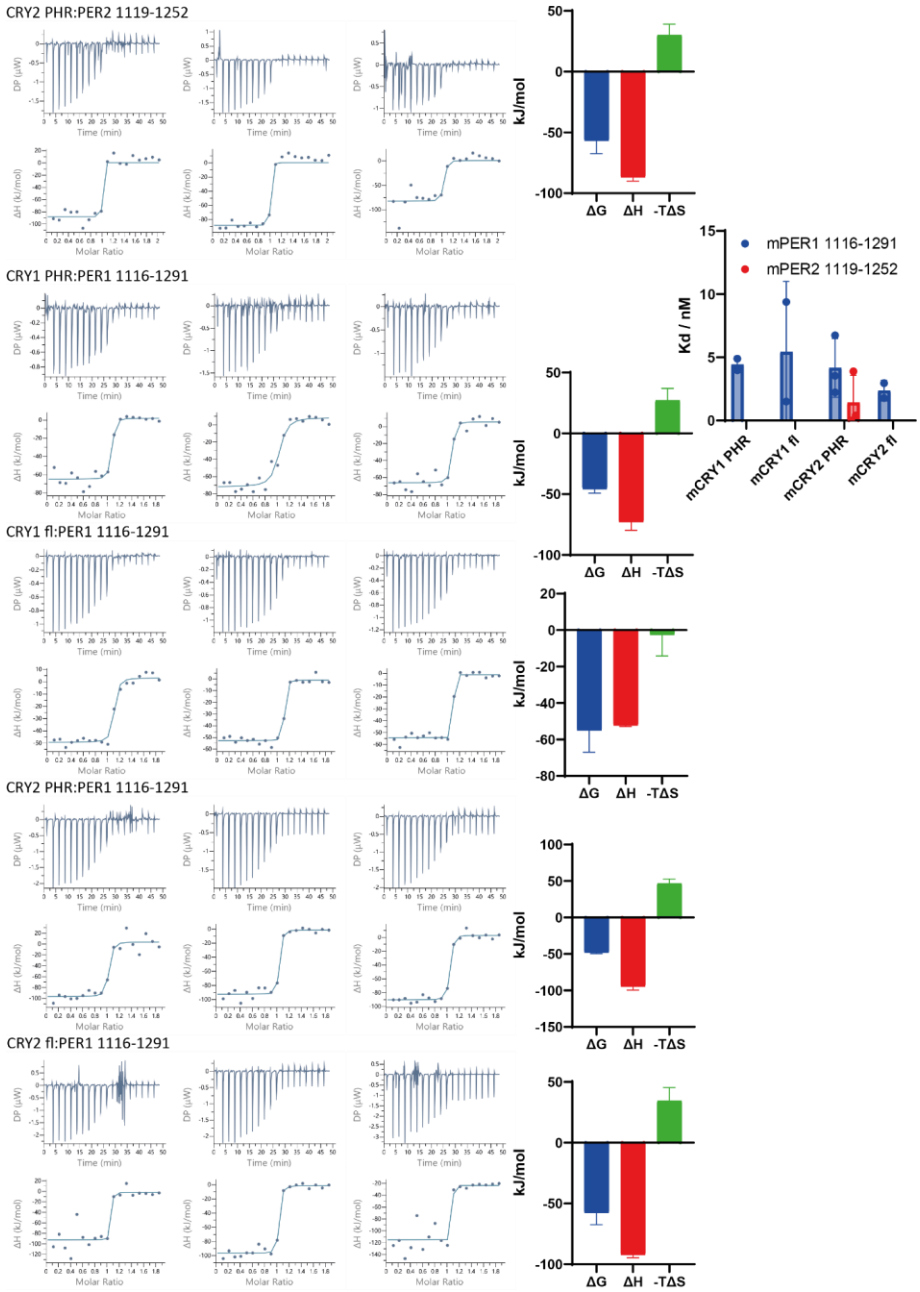

Figure 38: ITC result of CRY:PER interaction. CRY1/2 fl/PHR affinity to PER1 1116-1291 and PER2 1119-125. The replicate ITC profiles are shown on the left, the signature plots of the triplicates on the right.

The affinities are shown in the far right. The affinities are insignificantly different between CRY paralogs and fl/PHR for PER1$_{1116-1291}$. Only CRY2 PHR:PER$_{1119-1252}$ affinity was measured.

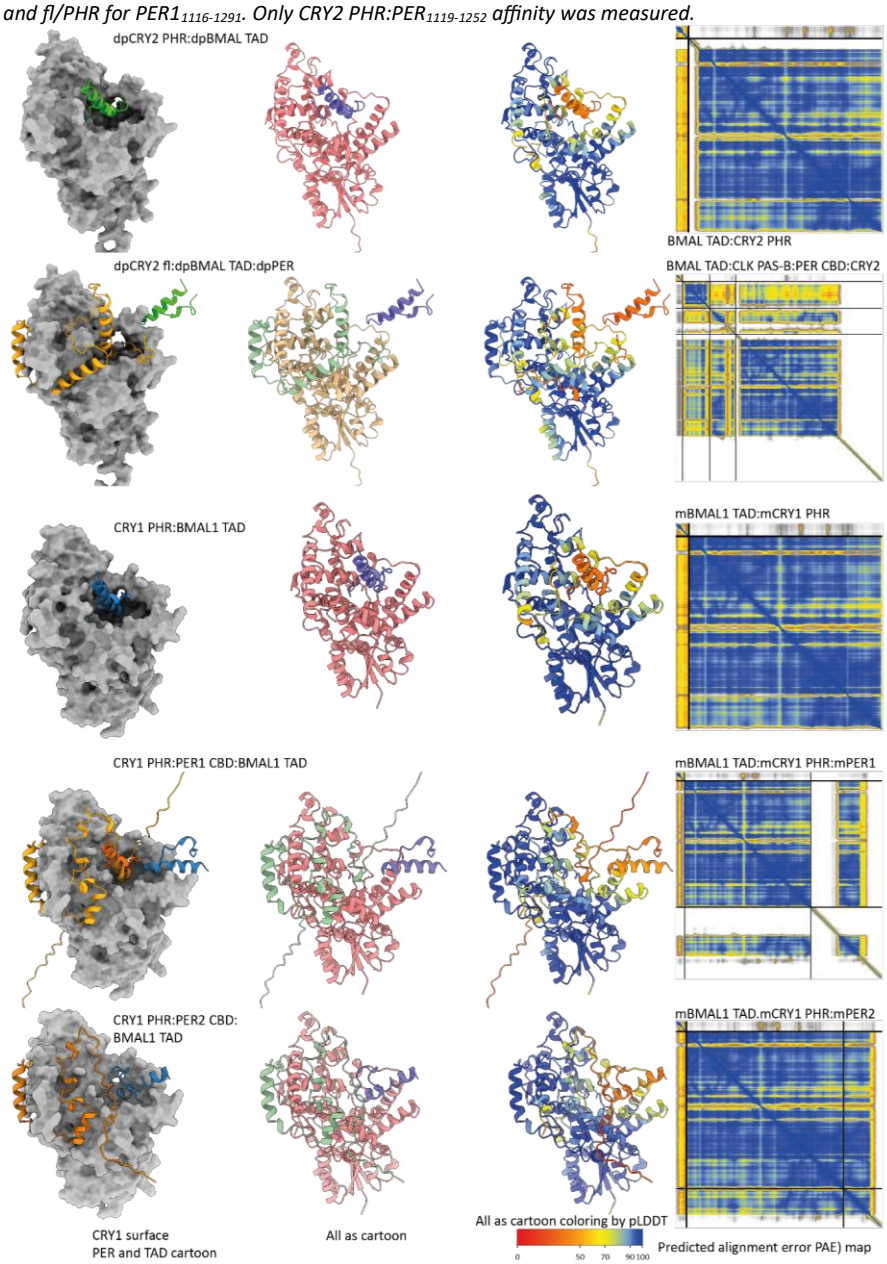

Figure 39: AlphaFold multimer predictions with prediction scores. On the left the representation used throughout the thesis is shown, second from the left a cartoon representation with chain coloring. The cartoon representation second from the right is colored according to pLDDT score with the color key

given at the bottom. Additionally, the PAE map is shown on the right. Interestingly the dpPER PAE trace has some higher confidence C-terminally of the also blue colored CBD. The corresponding region is at the bottom of the primary pocket and has a higher pLDDT than its surrounding PER residues.

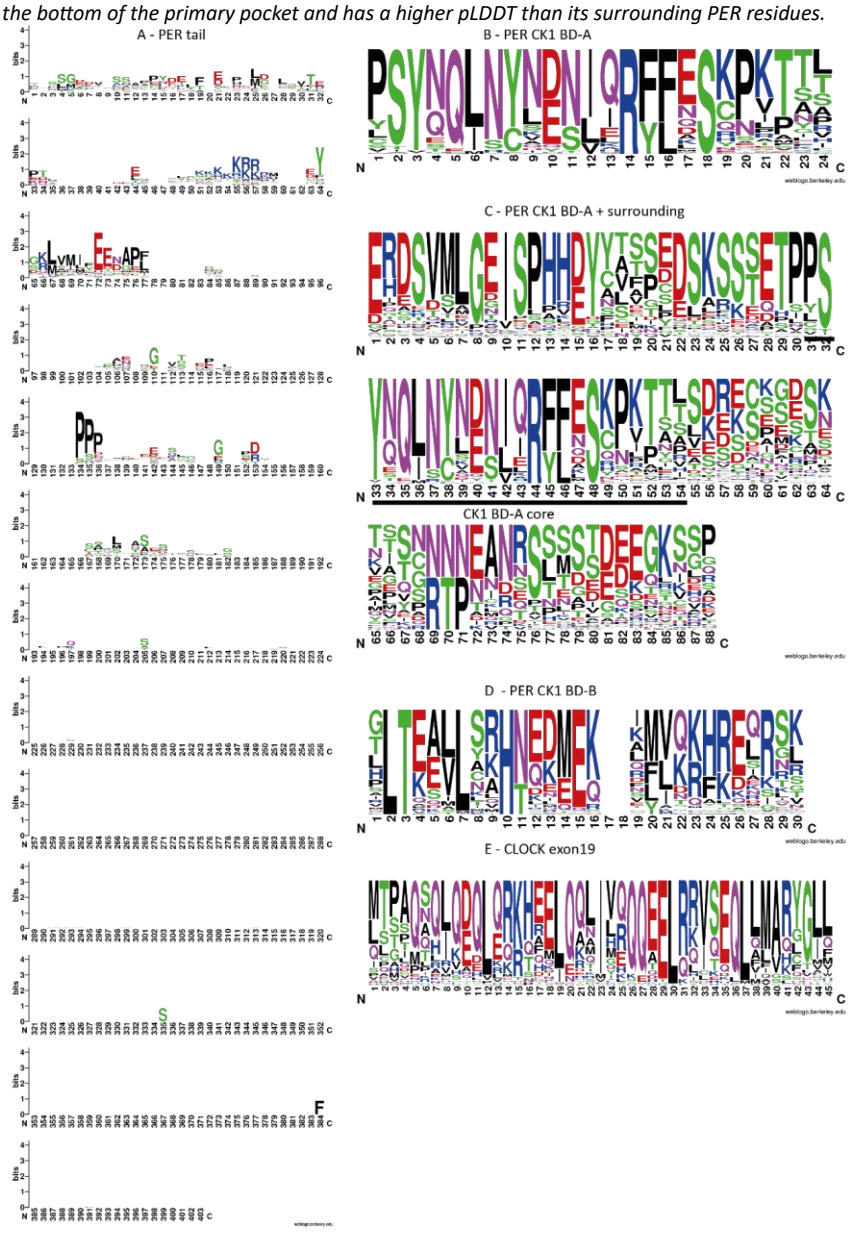

Figure 40: Sequence logo of PER and CLOCK motifs . A: Logo of the PER tail from all PERs in this study. B: Logo of the CK1 BD-A and C: logo of CK1 BD-A with the surrounding partially conserved amino acids. D: Logo of CK1BD-B. E: logo of CLOCK exon19

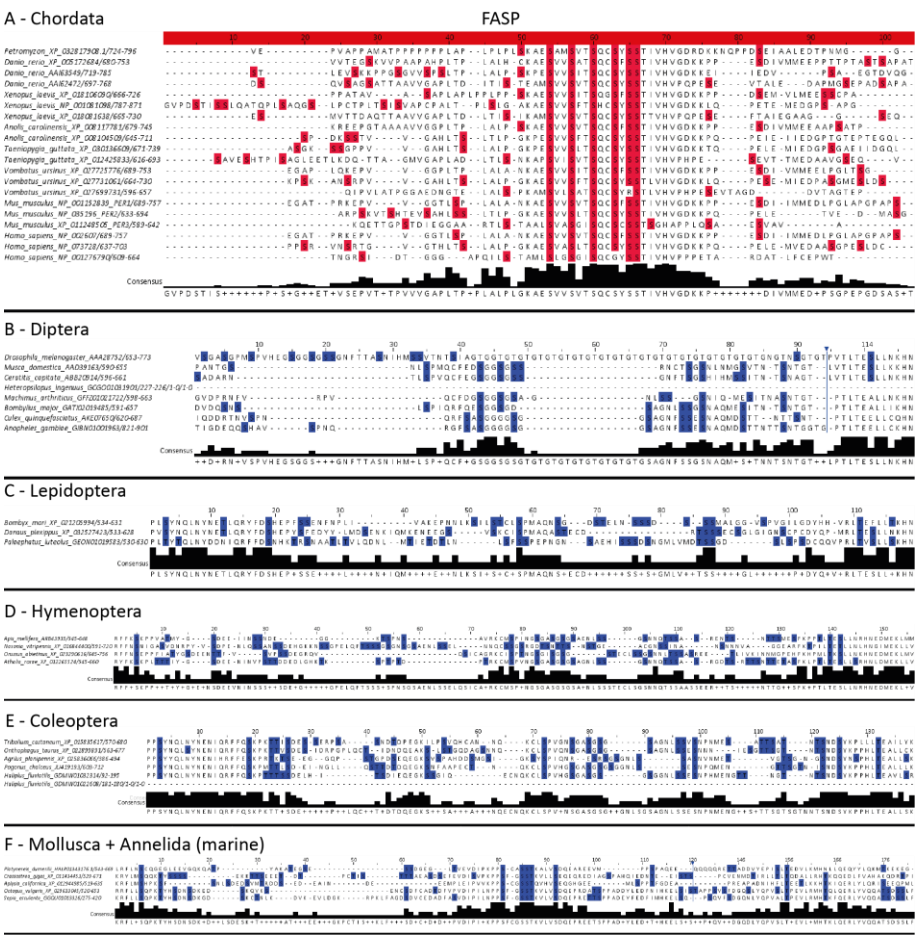

Figure 41: The FASP region for insect orders, Chordata and marine organisms. The mammalian FASP region (red) in an alignment of Chordata PER and potential phosphorylation sites (blue) for some insect orders and the marine organismsthat have a PER are shown.

Acknowledgements

I would like to sincerely thank all the people who supported me in one way or another along my Ph.D./Dr. journey.

I would like to thank my supervisor Prof. Eva Wolf for giving me the opportunity to be part of her group, work on this interesting and challenging project and independently develop and test my ideas. Our fruitful discussions, her constructive feedback and advice helped me to develop my ideas and led to the successful competition of this thesis.

I would also like to thank the members of my thesis advisory committee, Prof. Miquel Andrade, Prof. Susanne Foitzik and Dr. Gilles Travé for their support and critical feedback on my work. I am greatly thank ful to Prof. Helen May-Simera as second examiner of this thesis.

I thank all the past and present members of the AG Wolf, … , … , …, …, …, …, …, …, …, …, …, … and … for the scientific discussions, help with lab and administrative work and for creating such a nice and collaborative working environment.

Special gratitude goes to my collaborators … (Pharmaceutical and Biomedical Sciences, Uni Mainz) for peptide synthesis and labeling, … (PCF, IMB) for XL-MS measurements and analysis, … and … (Charité Berlin) for reporter gene assays, … and … for bioinformatics support, helpful discussions and explanations. Their support enabled this thesis in the first place and extended my method range considerably.

I want to extend my gratitude to … for AlphaFold installation and support, … for support on RNA isolation and cDNA synthesis, IMB media lab for media production, … and … from the IMB protein production core facility for scientific discussions, help and the wonderful enzymes they produce for all of IMB.

It was my pleasure to work with the talented young scientists Jule Urschel, Til Wanner and Verena Schneider of which the first two did their bachelor theses and Verena an internship. It barely felt like supervision, it was fun and I hope the best for your lifes and theses and careers!

My special thanks goes to … and … for being great and patient mentors.

This thesis was only possible due to generous funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) to the GRK2526/1 – Projectnr. 407023052 (GenEvo) that enables not only my work but of all GenEvo students. And this GenEvo community enriched my life during and even more after the Covid-19 pandemic. These fantastic people and talented scientists are …. . This of course was only possible by keeping all of Ph.D. students and PIs us organized, an enduring and crazy task that

our coordinators … and … took up. Thank you very much … and …! Finally, this program is formed and organized by our spokespersons … and … thank you for making GenEvo a great interactive and growing network allowing us young scientists to work on our thesis with all necessary resources and training.

Nothing of this would have been possible without my family and friends. I want to thank … for always being there and suggesting to kick-off this writing process in his flat. I want to … for being such a wonderful and great person. Many thanks to my non academia friends …, … and … for their constant support.

My highest thanks go to …, 10 Bresso … and …, … and … for being great friends and fantastic climbing partners into whose hands I can literally place my life anytime. Climbing up some rock face is the best way to forgot just about everything, even a thesis and inconclusive experiments. To many more adventurers high above!

My fastest thank goes out to …, …, …. and … for running together and representing our groups at the company fun runs, Mainz Marathon and motivating each other to keep on running (faster and faster).

Special thanks go to my family, who always supported me and especially to my mom who made my studies possible and motivated me to do a Ph.D.

Last but not least I want to thank … because she is always there for me and because I love her.

CV