Modelling the Circadian Clock

A thesis submitted to the University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Biology, Medicine and Health

2023

Alex A. Koch

School of Medical Sciences
Division of Diabetes, Endocrinology and Gastroenterology

Contents

Word Count: 88545

List of figures

List of publications

Alex A Koch, James S Bagnall, Nicola J Smyllie, Nicola Begley, Antony D Adamson, Jennifer L Fribourgh, David G Spiller, Qing-Jun Meng, Carrie L Partch, Korbinian Strimmer, Thomas A House, Michael H Hastings, Andrew SI Loudon. (2022) Quantification of protein abundance and interaction defines a mechanism for operation of the circadian clock. *eLife* 11:e73976. doi.org/10.7554/eLife.73976

Nicola J. Smyllie, James Bagnall, **Alex A. Koch**, Dhevahi Niranjan, Lenka Polidarov, Johanna E. Chesham, Jason W. Chin, Carrie L. Partch, Andrew S. I. Loudon, and Michael H. Hastings. (2021) Cryptochrome proteins regulate the circadian intracellular behavior and localization of PER2 in mouse suprachiasmatic nucleus neurons. *PNAS* doi.org/10.1073/pnas.2113845119

Terms and abbreviations

Throughout this thesis the names of genes are italicised and have their first letter capitalised (*Period2*) whereas proteins are all capitalised (i.e. PERIOD2). Two colons are used to denote fusions of two proteins (i.e. PERIOD2::LUCIFERASE) while protein-protein binding is represented by a single colon (i.e. CLOCK:BMAL1). Species names of animals and organisms are italicised and have their first letter capitalised (i.e. *Drosophila*). To distinguish between genes originating from different species but with the same name, a preceding letter is placed before the name of the gene, specifically 'm' for *Mus musculus* (i.e. *mPer2*), 'h' for *Homo sapiens* (i.e. *hPer2*), and 'd' for *Drosophila* (i.e. *dPer2*). Concentrations of molecular species are represented by square brackets to avoid confusion between complexes and their constituents. For example, concentrations of the monomeric proteins CLOCK and BMAL1 are [CLOCK] and [BMAL1] respectively, which may form the dimer CLOCK:BMAL1 with a concentration of [CLOCK:BMAL1]. Unless otherwise stated all concentrations are in nanomolar, nM, and diffusion rates are in micrometers squared per second, $\mu m^2 s^{-1}$.

AIC Akaike information criterion
ALF Adult lung fibroblast
ARNTL Aryl hydrocarbon receptor nuclear translocator-like protein 1
bHLH Basic helix-loop-helix
BFP Blue fluorescent protein
BMAL1 Brain and muscle ARNT-like protein 1
CAS CRISPR associated protein
CCGs Clock controlled genes
ChIP Chromatin immunoprecipitation
CI Confidence interval
CK1 Casein kinase 1
CLOCK Circadian Locomotor Output Cycles Kaput
CRISPR Clustered regularly interspaced short palindromic repeats
CRY Cryptochrome Protein
CT Circadian time
D-box D-box cis-regulatory element DNA site. The common sequence of a D-box is TTATG(C/T)AA
DBD DNA binding Domain
DBP Albumin gene D-site binding protein
DEX Dexamethasone
DMSO Dimethyl sulfoxide
DNA Deoxyribonucleic acid
E-box Enhancer box DNA cis-regulatory element. Consensus sequence is CANNTG.
EGFP Enhanced green fluorescent protein
FBS Fetal bovine serum
FCS Fluorescence correlation spectroscopy
FCCS Fluorescence cross-correlation spectroscopy
FP Fluorescent protein
FRAP Fluorescence recovery after photo-bleaching

GAPDH Glyceraldehyde 3-phosphate dehydrogenase
HEK Human embryonic kidney
Hygro Hygromycin
ipRGCs Intrinsically photosensitive retinal ganglion cells
KI Knock in **KO** Knock out
LUC Luciferase
LV Lentivirus
MCMC Markov chain Monte Carlo
MEF Mouse embryonic fibroblast
MLE Maximum likelihood estimation
mRNA messenger ribonucleic acid
NEAA Non-essential amino acids
Neo Neomycin
NF-κB Nuclear factor kappa-light-chain-enhancer of activated B cells
NPAS2 Neuronal domain containing protein 2
ODE Ordinary differential equation
PAS Per-Arnt-Sim
PBS Phosphate buffered saline
Rev-Erb Reverse-ERB
PCR Polymerase chain reaction
PER Period protein
PFA Paraformaldehyde
PMT Photo multiplier tube
PPI Protein-protein interaction
Puro Puromycin
RT-qPCR Reverse transcription quantitative polymerase chain reaction
RFP Red fluorescent protein
RNA Ribonucleic acid
ROR Retinoic acid receptor related orphan receptor
RORE Retinoic acid orphan nuclear receptor response element
SCN Suprachiasmatic nucleus
SEM Standard error of the mean
smFISH Single molecule in-situ hybridisation
SSA Stochastic simulation algorithm
SD Standard deviation
SF Skin fibroblast
TAD Transactivation domain
TF Transcription factor
TIM Timeless protein
TTFL Transcription Translation Feedback loop
UBC Promoter of polyubiquitin C gene
Var Variance
VEH Vehicle
WT Wild-type
ZT Zeitgeber Time

Abstract

Organisms have evolved molecular clocks to track 24-hour day/night light cycles which regulate all important physiological functions including metabolism, immunity, and cell-cycle. Cell-autonomous molecular clocks operate as a transcription-translation feedback loop (TTFL) driven by CLOCK:BMAL1, driving rhythmic transcription of genes including the negative repressors *Cry* and *Per*. Delayed negative feedback is initiated by CRY1 directly inhibiting BMAL1, followed by PER1/2 mediating removal of CLOCK:BMAL1 from target E-box DNA sites. Thus, protein-protein interactions are the key rate limiting steps which ensure correct phasing and periodicity. Previous studies have shown that the circadian clock is driven by impressively low abundance mRNA and protein copy number, but how the clock robustly generates cycles with low molecular numbers is unclear. The current TTFL model is based largely on qualitative genetic-based studies and lacks quantitative validation of how specific components of the cellular clockwork interact in time and space within the cell.

I investigated how a finite pool of BMAL1 proteins regulate thousands of target sites over 24 hours. From this, I developed a quantitative model of CLOCK:BMAL1 binding DNA using single-cell data for fluorescent fusion protein dynamics and interactions measured using live-cell microscopy and Fluorescence Correlation Spectroscopy (FCS). I found that the approximately 1000 CLOCK:BMAL1 complexes are highly mobile, likely rapidly moving between the far greater number of DNA target sites. Modelling showed that CRY1 complex formation with PER2 regulates the DNA residence time, and that these PER:CRY complexes play a dual role as both transcriptional repressors and enhancers of CLOCK:BMAL1 mobility. This mechanism therefore allows low copy-number clock proteins to regulate a wide repertoire of thousands of gene targets.

Despite the clear importance of interactions in the circadian circuit, ex vivo measurements of the affinities of protein-protein interaction are lacking. I therefore undertook a study to measure these for all major components of the molecular clock. This demonstrated that many of the protein-protein interactions in the circadian clock are remarkably strong and well conserved over multiple cell-types and different protein concentrations. Combining these data with modelling, I found that most interactions were direct, without the need for additional facilitating partner proteins, showing the clock operates as a set of serial pairwise interactions.

To facilitate the use of FCS, I developed new analytical tools which improve accuracy and robustness of fit. I derived an approximate likelihood model and applied maximum likelihood estimation to directly analyse raw unprocessed FCS data to increase the available data density by three orders of magnitude. This new methodology can infer concentrations and diffusion rates with as little as a few milliseconds of data rather than the current several seconds.

Lastly, I present a new application, Network Designer, to enable quicker exploration of models by graphically constructing networks with automated generation of differential and stochastic equations. I used this software throughout this body of work to create mathematical models of the circadian clock.

By combining quantitative experiments and modelling of circadian proteins, I have offered new insights into when and how protein-protein and protein-DNA interactions may define the operation and generation of circadian rhythms in mammalian cells.

Declaration of originality

I hereby confirm that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

Dedication

This thesis is dedicated to my friends and family that have supported me throughout this incredible journey. Your love, encouragement, and unwavering belief in my abilities have carried me through the most challenging moments.

To my parents, Nicola and Dieter, I am deeply grateful for believing in me and trying your best to understand my many ramblings on science. Your unwavering support and unconditional love have been the foundation upon which I have built my success.

To my brother Christian, thank you for (literally) cheering me on, and asking insightful questions that make me think about the world more deeply.

To my grandparents, Tom, Sylvia, Siegrid, and Klaus, thank you for all of your love and understanding. Without your support I would not be where I am now.

To my friends, Alex, Andrew, Zak, and Josh, thank you for all the chats, games, and laughs that brightened my days and got me through my hardest times.

To Derya, and Ethan, thank you for your friendship, all of the crazy unforgettable experiences and getting this PhD over the line.

Finally, to my housemate and friend Bailey, I could not have done it without you. Thank you for listening to my rants, as well as supporting me through the best and worst times.

This accomplishment is not solely my own, but a testament to love, friendship, support, and guidance you have provided throughout the years. This thesis is for you, and I am eternally grateful to have you in my life.

"All We Have to Decide is What To Do With the Time That is Given Us"

- Gandalf in The Fellowship of the Ring by J. R. R. Tolkien

Acknowledgements

I would like to express my profound gratitude to a number of individuals who have contributed immensely to the success of this thesis. First and foremost, I am deeply grateful to my academic supervisors, Prof. Andrew Loudon, Dr. Thomas House, and Prof. Korbinian Strimmer, for their invaluable guidance, continuous support, and constructive feedback throughout the course of my PhD journey. Their expertise, enthusiasm, and encouragement have been instrumental in shaping the direction and quality of this research.

I would like to extend my appreciation to the QBB PhD programme, led by Prof. Nancy Papalopulu, for providing me with the opportunity to be part of an inspiring and stimulating research environment in biological sciences after my Master's in physics. I am particularly grateful to my fellow PhD students in the programme, whose passion for science and collaborative spirit have enriched my time at Manchester. It has also been a pleasure hear Nissirn Alachkar's insights and work closely with her on RNA imaging.

My gratitude extends to Dr. Jean-Michel Fustin for his expert advice on cell culture techniques, which significantly contributed to the development of the experimental aspects of this thesis. I am indebted to him for his time, patience, and willingness to share his knowledge.

I would like to express my heartfelt gratitude to the members of the Loudon lab group for their assistance and support during my time in the lab. I am particularly thankful to Nicola Begley, whose friendship, and willingness to help have been invaluable to the successful completion of this work. Special thanks go to James Bagnall for his guidance, patience, and friendship throughout my PhD. James' expert knowledge and insight was integral to the completion of my PhD.

Additionally, I would like to thank the Bioimaging facility at the University of Manchester, especially Dave Spiller, for all of their support, illuminating discussions, and teaching me how to use a microscope.

I would also wish to thank my long time housemate and friend Bailey Winstanley for not only keeping me sane, but for swiftly correcting my mathematics.

Lastly, I would like to acknowledge that this work was generously funded by the Wellcome Trust as part of the PhD programme in Quantitative and Biophysical Biology programme (Grant number 216416/Z/19/Z).

Chapter 1

Introduction

1.1 The circadian clock

1.1.1 What is the clock?

Darwin (1859) first described in 'The Origin of Species' that natural selection is the pressure placed on flora and fauna to adapt to changes in their environment. The rotation of our planet is no exception, driving 24-hour cycles of available sunlight and temperature, dramatically altering the conditions for life to flourish. Most organisms have adapted and thrived to these evolutionary pressures by evolving biological mechanisms to track and adapt to these 24-hour cycles (Pittendrigh, 1993). Jean Jacques d'Ortous de Mairan made the first recorded observations of circadian (Latin for 'about a day') rhythms in 1729 by noting the continued folding/unfolding of the leaves of a mimosa plant

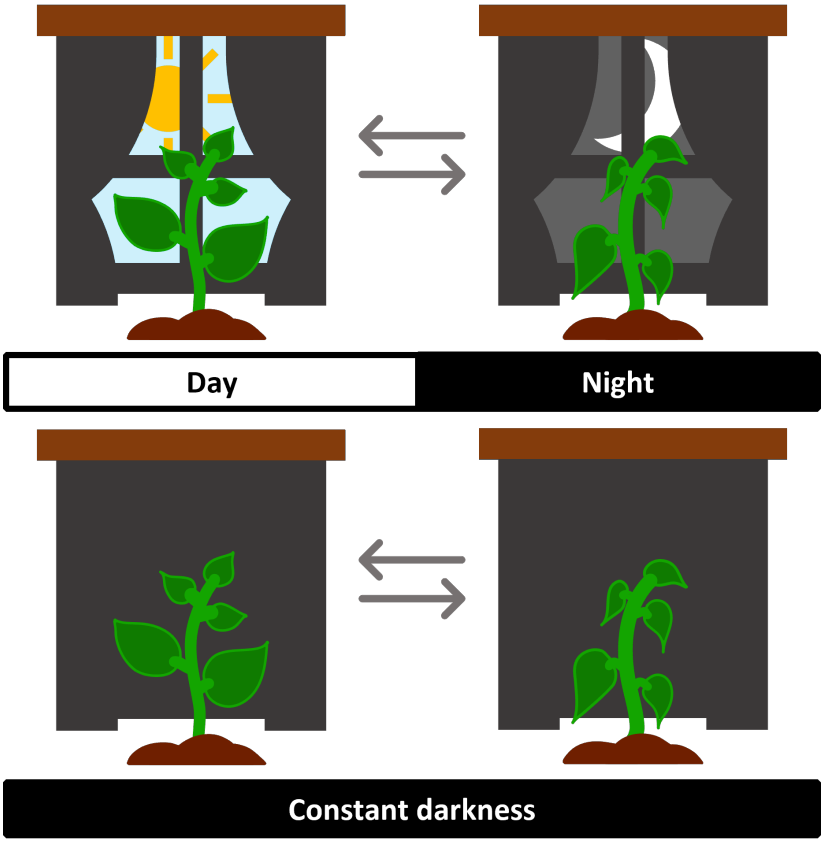

Figure 1.1: Jean Jacques d'Ortous de Mairan noted that the leaves of a mimosa plant unfolded in the morning and folded at night. When placed in a dark room the leaves continued this circadian rhythm for several days suggesting this is an innate feature of the cells within the plant. Adapted from Nobel Prize Outreach (2023).

placed in a dark room following normal daily light exposure (Kreitzman and Foster, 2011), see Figure 1.1. It has since been shown that plants regulate their metabolism over the day to optimise energy production given available sunlight, and animals modify their behaviour such as sleep/wake timings as well as locomotor activity, in the case of diurnal and nocturnal animals, to anticipate accessible food sources and energetic requirements (Wang et al., 2015). Are these clocks exogenous, simply tracking daily variations, or are they endogenous, capable of entrainment and anticipation of daily light rhythms? After observing cycles in his body-temperature, Aschoff (1960) showed through constant light exposure that circadian clocks are endogenous oscillators that 'free run' when all timing cues, or Zeitgebers, are removed. Persistent rhythms in the absence of external cues are a key feature of circadian clocks. Additionally, entrainment to the environment, often occurring in a phase dependent manner, closely mimics the synchronisation of a simple pendulum to external vibrations.

Organisms have evolved molecular circadian clockwork, common to almost all cells, which are built upon cycles of proteins repressing their own production and activity in a delayed manner. Simple single cell organisms such as *Cyanobacteria* as well as multi-cellular plants directly entrain cellular oscillators with day/night cycles (Ahmad et al., 1998; Leypunskiy et al., 2017). In contrast, mammals make use of a central pacemaker in the brain known as the Suprachiasmatic Nucleus (SCN). The SCN receives light input via the retina through intrinsically photosensitive retinal ganglion cells (ipRGCs) to synchronise a group of approximately 20,000 coupled oscillatory cells, which further distribute timing cues to downstream peripheral oscillators throughout the body via neuroendocrine cues (Meyer-Bernstein et al., 1999; Buijs et al., 2003). Myriad studies have linked the circadian clock to innate, physiological functions across numerous organisms, with mammals exhibiting daily rhythms in body-temperature (Buhr et al., 2010), cell-cycle (Matsuo et al., 2003; Bieler et al., 2014), hormones (Lightman et al., 2016), metabolism (Green et al., 2008), and immunity (Scheiermann et al., 2018).

The discovery of genetic circuits driving the circadian clock mechanism lead to the award of the Nobel Prize in Physiology or Medicine in 2017 to Jeffrey C. Hall, Michael Rosbash and Michael W. Young (Bargiello et al., 1984; Zehring et al., 1984). Despite significant advances in the genetic understanding of the circadian clock, quantitative knowledge is lacking on how this network of molecular interactions generates robust 24-hour cycles.

1.1.2 Discovery of the endogenous molecular clock

Circadian components were first identified via genetics

How do endogenous circadian rhythms, that free run in the absence of external stimuli, arise? Following studies in the fruit fly *Drosophila* that demonstrated that single genes could control physical traits, Benzer hypothesised that there may be genes that control circadian behaviour in flies. Konopka and Benzer went on to use a mutagenic screen on *Drosophila* to discover the first circadian gene, *Period*, in 1971, which when modified shortened (19 h), lengthened (28 h), or introduced arrhythmia to locomotor activity as well as eclosion (Konopka and Benzer, 1971). Despite this breakthrough it took until the late 80s for Hall and Rosbash to reveal a rhythm in the amount of dPER protein, detected using antibodies, in the visual system of *Drosophila* (Siwicki et al., 1988), leading to the subsequent discovery that *dPer* mRNA transcripts cycle with a peak a few hours prior to a peak in dPER protein (Hardin et al., 1990). Additionally, overexpression of dPER supresses *dPer* mRNA transcription, a discovery which lead to the development of what is now known as a transcription-translation feedback loop (TTFL) model of the clock. In this, transcription factor proteins rhythmically feedback to repress their own transcription (Zeng et al., 1994), resulting in the generation of rhythmic mRNA and protein. In *Drosophila*, dTIM (gene *Timeless*) was identified and found to oscillate as well as gate nuclear entry of PER providing a crucial mechanism required for generating 24 h cycles (Gekakis et al., 1995; Myers et al., 1995; Sehgal et al., 1995).

In the 1990s the first mammalian clock gene, *mClock* (Circadian Locomotor Output Cycles Ka-

put), was found by Takahashi's group in mice. *Clock* was discovered via mutagenesis screening for disrupted circadian phenotypes and subsequent positional cloning (Vitaterna et al., 1994; Antoch et al., 1997). This was followed shortly by the discovery of Brain and Muscle ARNT-Like 1 (*Bmal1*) independently by Hogenesch and colleagues and Ikeda and Nomura (Hogenesch et al., 1997; Ikeda and Nomura, 1997). Shortly after, biochemical studies showed that the protein products of these genes, CLOCK and BMAL1, were found to partner with one another to regulate the transcription of PERIOD1 protein via 6 base-pair DNA sequences, CANNTG, termed Enhancer Boxes (E-Box) which are present in promoter of the *mPer1* gene (Gekakis et al., 1998). Later, whilst searching for proteins responsible for environmental light sensing in the clock, van der Horst and colleagues identified a role for CRYPTOCHROME proteins. These proteins in plants serve as photolyases, and are important in DNA repair. Genetic studies in mice showed that when knocked-out independently, *Cryptochrome1* and *Cryptochrome2* (*mCry1* and *mCry2*) accelerated (22.5 h) and delayed (24.6 h) the clock respectively, thus demonstrating an antagonistic relationship between the two which is critical to proper timekeeping (van der Horst et al., 1999). Crucially knocking-out both alleles of *mCry* gene rendered mice arrhythmic under constant dark conditions. Interestingly, mice still retained the ability to entrain their locomotor activity when under standard light-dark cycles, suggesting CRYs are not light-sensing and in mammals have been adopted as essential core clock components regulating timing. Melanopsin expressed in the intrinsically photosensitive retinal ganglion cells (ipRGCs) was later shown to pass light cues to the SCN independent of rods and cones (Freedman et al., 1999; Hattar et al., 2002), providing further essential insight into how molecular re-setting of the circadian clock by light/dark cycles is achieved.

Multiple subsequent studies in mammals have identified additional ancillary interconnected loops that modify TTFL timing and output. For example, Preitner et al. (2002) demonstrated that *mBmal1* mRNA and protein oscillations are due to repression of orphan nuclear receptor REV-ERBα binding to RORE sites in the promoter of the *mBmal1* gene. REV-EBRα was further shown to be itself negatively regulated by PER/CRY, elucidating how *mBmal1* oscillates in anti-phase to *mPer1* and *mPer2* (Preitner et al., 2002). Discovery of these additional loops working in conjunction with the core TTFL has thrown important light into how tissue specific repertoires of circadian timing are achieved.

How can cells generate 24-hour cycles?

A dominant model in our thinking of how biological rhythms are generated is the concept of delayed negative feedback. In this, robust circadian oscillations are generated by the expression of repressive components (PER, CRY and REV-ERB in mammals) which exhibit delayed feedback to repress transcription (Gonze and Ruoff, 2021). A clear prediction is that tuning of the life-expectancy of these repressive elements will have profound effects on period generation, and here, protein degrading kinase enzymes play a crucial role. Transcription and translation are biologically rapid processes, on a time scale of seconds to minutes (Shamir et al., 2016). Thus, the pace of the clockwork is set by the regulation of protein metabolism in the cell. An early candidate for this emerged in *Drosophila* with the discovery that Casein Kinase 1 (CK1, encoded in the *Doubletime* gene) plays a key in the regulation of PERIOD protein degradation of PER through phosphorylation, leading to ubiquitination and proteasomal degradation. Mutations in *Doubletime* have a profound impact on the *Drosophila* clock. Mammals possess a closely related CK1 kinase which is also implicated in regulating mammalian PER proteins (Kloss et al., 1998; Xu et al., 2005; Keesler et al., 2000; Narasimamurthy and Virshup, 2021). A gain of function mutation of CK1ϵ, known as tau in mice and Syrian hamsters, is responsible for the significant acceleration of rhythms from 24 to 20 hours, demonstrating the important rate setting role of kinases in the clock (Ralph and Menaker, 1988; Lowrey et al., 2000; Meng et al., 2008). Subsequently, mutations of the CK1ϵ binding domain of human PER2 leads to perturbed rhythms and a striking 4-hour advance of sleep onset in the case of familial advanced sleep phase syndrome (FASPS) (Toh et al., 2001). In all cases, phosphorylation of

PER2 is perturbed, leading to increased degradation by the ubiquitin–proteasome system.

It is hypothesised that post-translational modifications, including phosphorylation by kinases, define an evolutionarily conserved mechanism for circadian clock function across species regardless of specific separately evolved TTFL components (Wong and O'Neill, 2018). Different organisms have evolved diverse components underlying the core mechanisms driving circadian clocks, with convergent evolution leading to multiple homologues of the molecular clock (Young and Kay, 2001). In addition to mammals and *Drosophila*, Clocks in *Cyanobacteria* (Johnson et al., 1996), *Neurospora* (Merrow et al., 2001), monarch butterflies (Froy et al., 2003), have all been extensively studied. The clocks of all of these organisms operate on the same principal of oscillating biochemical reactions (Young and Kay, 2001). Initially, it was thought that prokaryotes like *Cyanobacteria* are too simple to support a clock as they possess a genome of only 4.5 million base pairs (Alvarenga et al., 2017). Hence, instead of a clock based on transcription as found in eukaryotes, *Cyanobacteria* use a mechanism of cyclic protein phosphorylation of the key KaiC protein. Strikingly, the *Cyanobacteria* clock has been shown to oscillate in vitro, requiring only the presence of Kai proteins and ATP, with no additional cell machinery required (Nakajima et al., 2005). Crucially, even through protein dynamics operate on a sub-second timescale, a 24-hour period is produced by the slow ATPase activity of KaiC. The rate of this catalytic process is closely tied to the period of the cyanobacterial clock (Abe et al., 2015). This underscores the importance of delays in establishing 24-hour circadian rhythms, especially when assembled from components that have naturally fast dynamics.

This raises the important questions as to what the key components of the mammalian clock are, and how they couple with other cellular systems to confer timing information.

1.1.3 Time conventions

Conventionally, the intrinsic time of circadian oscillators, whether they are cells or whole animals, is defined by circadian time, abbreviated to CT. CT references the time of an organism's endogenous circadian clock without regard to environmental cues or other entraining cues, otherwise known as zeitgebers (German for 'time giver') (Xie et al., 2019). CT0 defines the onset of activity in diurnal animals and CT12 marks the onset of activity in nocturnal animals. CT is periodic, so CT24 is equivalent to CT0 for organisms with an intrinsic circadian period of 24 hours (Karatsoreos and Silver, 2017). Zeitgeber time (ZT) is useful in defining the start of a zeitgeber which represents an external time cue used to perturb circadian rhythms. For instance, in the context of light/dark cycles, ZT0 would signify when the lights turn on. It therefore follows that for organisms with a 24-hour period under regular 12:12 light/dark cycles, ZT and CT are equivalent to one another.

1.2 Mammalian clocks

1.2.1 Why should we study mammalian clocks?

Circadian clocks are a ubiquitous feature of many organisms and thus the study of clocks is wide and varied. Core elements found in mammalian clocks, such as *mClock*, *mBmal1*, and *mPer*, have the orthologs *dClock*, *dCycle* (*Cyc*), and *dPer* respectively in *Drosophila*. However, in *Drosophila* the key partner for PER is TIMELESS (TIM) rather than CRY, which whilst TIM is present, it is not essential for rhythm generation in mammalian clocks (Gotter et al., 2000). Additionally, in *Dosophila*, CRY is light sensitive and part of the light resetting response. In this thesis, I restrict my studies to the molecular workings of the mammalian clock, using the mouse *Mus musculus* as a model organism. Mammals share many similar features in how they generate and organise circadian rhythms across the body; a central pacemaker residing the brain known as the SCN receives timing information in the form of light impulses which are then subsequently passed to autonomous peripheral oscillators residing in the rest body's cells. Handily, many of the key elements comprising cell autonomous clocks are conserved across mammals (Young and Kay, 2001). Currently, there is

much interest in the medical treatment of circadian disorders through combinations of light therapy and pharmaceutical interventions such as the hormone melatonin (Cederroth et al., 2019). Beyond the direct treatment of circadian disorders, there is emerging evidence that emphasises the significance of circadian timing in pharmacological treatments. A variety of studies have indicated that the timing of drug dosing, especially in the context of cancer treatments, can lead to improved health outcomes. This chronopharmacological approach suggests that by aligning drug administration with specific times in the circadian cycle, tissues can be exposed to drugs when they are most able to tolerate toxicity, thereby maximising therapeutic efficacy and minimising side effects (Lévi et al., 2010; Annabelle Ballesta et al., 2017; Cederroth et al., 2019; Ruben et al., 2019). Thus, research findings made whilst studying mice could indicate new routes of research of clinical relevance which may be pursued to improve health outcomes in humans. Further characterisation of how the components of the circadian clock come together and interact to generate daily rhythms can elucidate effective targets for these interventions.

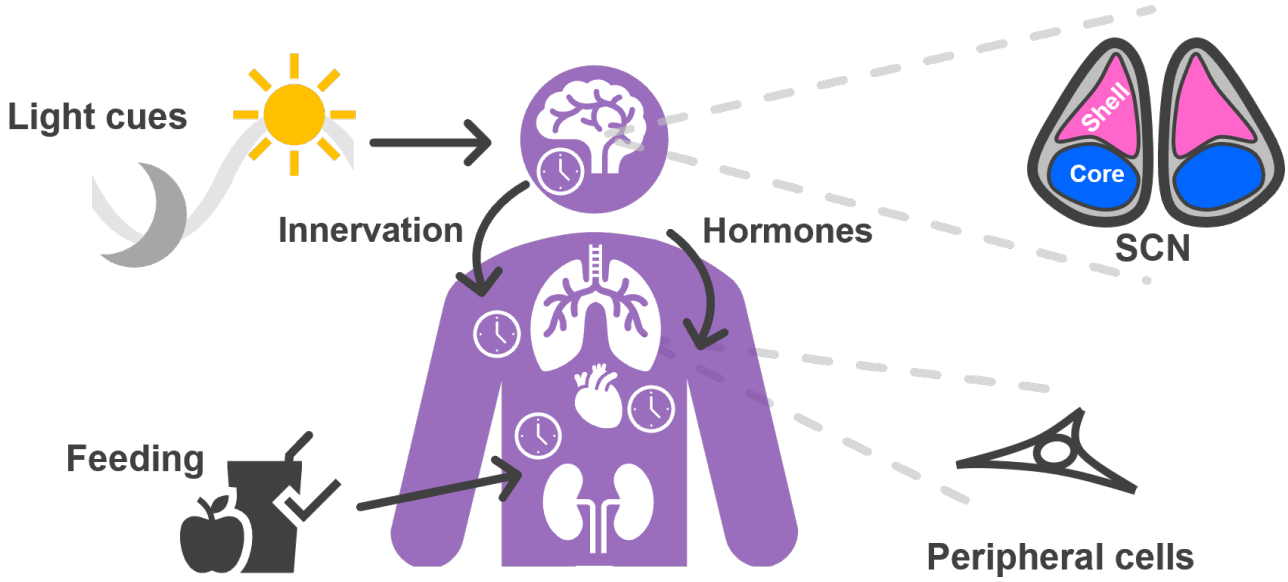

Figure 1.2: Light cues enter the eyes via the retina and are passed directly to the suprachiasmatic nucleus (SCN) in the brain which coordinates timing throughout the body. Entrainment of peripheral oscillators, such as fibroblast cells, is achieved by direct innervation, hormonal cues and body temperature. Additional entrainment by external cues such as feeding also occurs in the liver. Both peripheral tissues and the neurons comprising the SCN possess autonomous molecular oscillatory systems known as the transcription-translation feedback loop.

1.2.2 The Suprachiasmatic Nucleus is the central pacemaker

In mammals the circadian clock is known to be coordinated in a hierarchical manner, see Figure 1.2. A collection of approximately 20,000 neurons within the suprachiasmatic nucleus (SCN) in the ventral hypothalamus entrains peripheral autonomous oscillators in cells and tissues throughout the body. The SCN is synchronised to environmental light cues entering the eyes via the retina, where melanopsin expressing intrinsically photosensitive retinal ganglion cells (ipRGCs) detect light exposure and project directly from the eyes to the SCN. Without melanopsin, mice have been shown to have reduced photoentrainment ability (Panda et al., 2002b). Key output mechanisms synchronising peripheral tissues include direct autonomic innervation (van der Horst et al., 1999), and hormonal cues (i.e. glucocorticoids) (Kaneko et al., 1981; Meyer-Bernstein et al., 1999). But crucially, metabolic cues driven by rhythmic feeding are dominant, such that the impact of the SCN "brain-clock" can be bypassed by imposition of timed feeding (Bass and Lazar, 2016). Lesioning of the SCN in rats has been shown to abolish both entrainment to light and free running circadian rhythms in the hormone adrenal corticosterone (Moore and Eichler, 1972) as well as in locomotor activity and drinking (Stephan and Zucker, 1972). An early key demonstration of the dominant role of the SCN in regu-

lating behaviour came from studies where grafts of foetal donor SCNs into hamsters with previously lesioned SCNs rescued circadian wheel-running rhythms (Lehman et al., 1987). When SCN tissue from short period tau (τ) mutants was transplanted into wild type animals, the recipient exhibited circadian periods identical to the donor, 22 hours from heterozygous tau SCNs (τ/WT) or 20 hours for homozygous tau SCNs (τ/τ) (Ralph et al., 1990). Interestingly, transplant studies such as these are effective in rescuing behavioural activity rhythms, but ineffective in rescue of most other outputs, including endocrine, sleep and temperature rhythms. This shows that the mechanisms of entrainment by the SCN are complex and involve both humoral signals (behaviour) as well as "hard-wired" outputs. Thus, the SCN plays a crucial role in controlling circadian rhythms across the body.

Individual neurons within the SCN have been shown to possesses an autonomous TTFL oscillator whereby CLOCK:BMAL1 activates transcription of PER/CRY which in turn attenuates the transcriptional capability of CLOCK:BMAL1 (Hastings et al., 2018). The neurons of the SCN form a highly coupled system which is spatially organised into dorsal neurons in the 'shell' and ventral neurons at the 'core' (1.2). These autonomous oscillators, capable of differential cycling to one another, spatially synchronise by passing timing information between each other via γ-aminobutyric acid (GABA) in the dorsal shell and vasoactive intestinal polypeptide (VIP) in the core (Hastings et al., 2003). When this coupling is removed by dispersing neurons in vitro or treating with inhibitors such as tetrodotoxin (TTX) neurons lost cycling over days, as measured by the PER2::LUC integrated reporter gene (Yoo et al., 2004), whereas when allowed to couple in intact tissue explants, the neurons of the SCN can oscillate almost indefinitely for weeks and even months outside of the body in the right culture conditions (Webb et al., 2009). These experiments combined with modelling studies have shown that intercellular coupling between the individual weak neuronal oscillators aids in the generation of robust rhythms tissue-wide by buffering against molecular noise (Gonze and Goldbeter, 2006).

1.2.3 Peripheral oscillators

The role of the SCN is to synchronize the internal clockwork to the phase of the external light-dark cycle and individual cellular clocks are capable of self-sustained oscillations in culture without SCN input. A key early observation using fibroblast cells (Balsalobre et al., 2000) demonstrated that peripheral cellular clocks are as competent as neurons in terms of robust time-keeping, but that they are not coupled, as demonstrated by Welsh and colleagues in studies of PER2::LUC in dispersed fibroblasts cells. In this latter case, cells oscillated in culture for prolonged periods, sometimes up to months at a time (Welsh et al., 2004; Yoo et al., 2004).

Mammals possess multiple copies of some of their core clock genes. In each case, deletion of a single paralog in mammals of core the clock components *Per1*, *Per2*, *Cry1*, *Cry2*, *Nr1d1* (Rev-erbα) and *Nr1d2* (Rever-erbβ), and Clock causes disruption of period and robustness, but not complete loss of circadian-like oscillations. There is one exception, *Bmal1* (also known as Mop3). This gene has a paralog, *Bmal2*, but the two genes are not homologous in function. Targeted deletion of *Bmal1* causes loss of circadian rhythmicity in cells, and as such is the only member of the TTFL family in mammals which is not compensated for by another protein (Bunger et al., 2000; Ikeda et al., 2000). Deletion has profound health implications and in mice lacking *Bmal1* in the liver hepatocytes exhibiting hypoglycemia during fasting, increased hyperlipidaemia, and atherosclerosis (Lamia et al., 2008; Pan et al., 2016). In *Bmal1* deficient macrophages, there is a gain in phagocytic activity (Kitchen et al., 2020). Rhythmic circadian gene and protein expression occur throughout the body due to these internal oscillators within each cell, with approximately 10% of the transcriptome cycling. Constitutively expressed CLOCK (Lee et al., 2001) and its paralog neuronal PAS domain-containing protein 2 (NPAS2) both interact with BMAL1 to maintain circadian function (Huang et al., 2012), with NPAS2 compensating for CLOCK in its absence (DeBruyne et al., 2007).

The circadian clock exerts control over many cellular functions, with variable phase and strength. However, which mRNA and protein are rhythmically controlled by the clock changes from tissue to

tissue. Additionally, outside of the core clock components, multiple studies have found <1% match between which transcripts cycle in SCN, liver, heart, and skeletal muscle tissues (Panda et al., 2002a; Storch et al., 2002). These differences in clock-controlled outputs occurs despite each cell possessing the same molecular machinery. Hence, internal clocks within each organ have an essential role to play in proper cellular regulation and tissue function.

1.2.4 Molecular clock architecture

CLOCK:BMAL1 drives oscillations by binding DNA target sites

At the core of the mammalian TTFL, the heterodimeric transcription factor complex CLOCK:BMAL1 form part of the basic helix-loop-helix transcription factor family, which searches the genome to bind proximal consensus E-box DNA sites upstream of transcription starts sites (TSS) of many genes. These include *mPer1-3* and *mCry1-2* (Yoo et al., 2005; Fustin et al., 2009), see Figure 1.3. E-box binding by CLOCK:BMAL1, alongside recruitment of co-activators CREB-binding protein (CBP) and Histone acetyltransferase p300, primarily induces transcription of core clock genes *mPer1-3* and *mCry1-3* during the day in mice as well as several hundred clock-controlled genes (CCGs) differentially over the whole circadian 24-hour cycle (Takahashi, 2017). In addition to E-boxes, chromatin immuno-precipitation (ChIP) studies have revealed synchronous binding of CLOCK:BMAL1 to 2000-6000 (tissue dependent) target sites in genes (Rey et al., 2011; Koike et al., 2012; Chiou et al., 2016; Oishi et al., 2017; Wu et al., 2017; Dyar et al., 2018; Beytebiere et al., 2019). Binding amplitudes – globally and per gene – oscillate over the day, with gene to gene transcriptional output of CLOCK and BMAL1 targets being highly heterogenous (Trott and Menet, 2018), likely owing to gene specific co-factors and different partner proteins, for example Hypoxia-inducible factor 1-alpha (HIF1α) (Wu et al., 2017). DNA binding varies over the day as a function of CLOCK:BMAL1 abundance, chromatin accessibility, and the CLOCK:BMAL1 affinity for DNA. Intriguingly, quantitative mass spectrometry studies show CLOCK and BMAL1 cycles around 20,000 protein copies per cell, suggesting after dimerization that the number of available CLOCK:BMAL1 is likely less than the estimated number of target sites (Narumi et al., 2016). For this reason, CLOCK:BMAL1 is hypothesised to be highly mobile, moving from site to site in contrast with the stationary constitutively bound model proposed by Lee et al. (2001). This is consistent with imaging experiments measuring a short DNA residence time of $\sim$ 10 s on for CLOCK:BMAL1 on E-boxes within transfected concatemers of the D-Box Binding PAR BZIP Transcription Factor (*Dbp*) gene (Stratmann et al., 2012). Moreover, theoretical models and data place target DNA site residency times for transcription factors, such as GR, P53, and Stat1, in the range of few seconds to tens of seconds (Hettich and Gebhardt, 2018; Azpeitia and Wagner, 2020). CLOCK:BMAL1 binding and transactivation of repressors *Per1-3* and *Cry1-2* defines the start of the feedback loop, and following translation of their mRNA, resultant CRYPTOCHROME (CRY) and PERIOD (PER) proteins accumulate in the cytoplasm, get phosphorylated, bind to one another, and then translocate into the nucleus to inhibit their own transcription by repressing CLOCK:BMAL1 complexes (Takahashi, 2017). Importantly, as PERs and CRYs repress their own transcription through CLOCK:BMAL1 in a dose dependent manner, this requires precise control of protein copy number in each cell, as the repressive action of these proteins define the pace of the clock.

Post-translational processes set the pace of the clock

Predicted delays between *Per* mRNA levels and PER proteins were first observed in *Drosophila* using techniques such as RNA hybridisation and protein-detecting antibodies respectively, (Siwicki et al., 1988; Hardin et al., 1990). However, this masks the true underlying dynamics of the clock required for rhythmic feedback. Typically, protein levels rise in quick succession with mRNA as there is little delay added by RNA processing, nuclear export, and translation. It is also likely that the phase difference between mRNA and protein for core clock components is due to complex post-translational

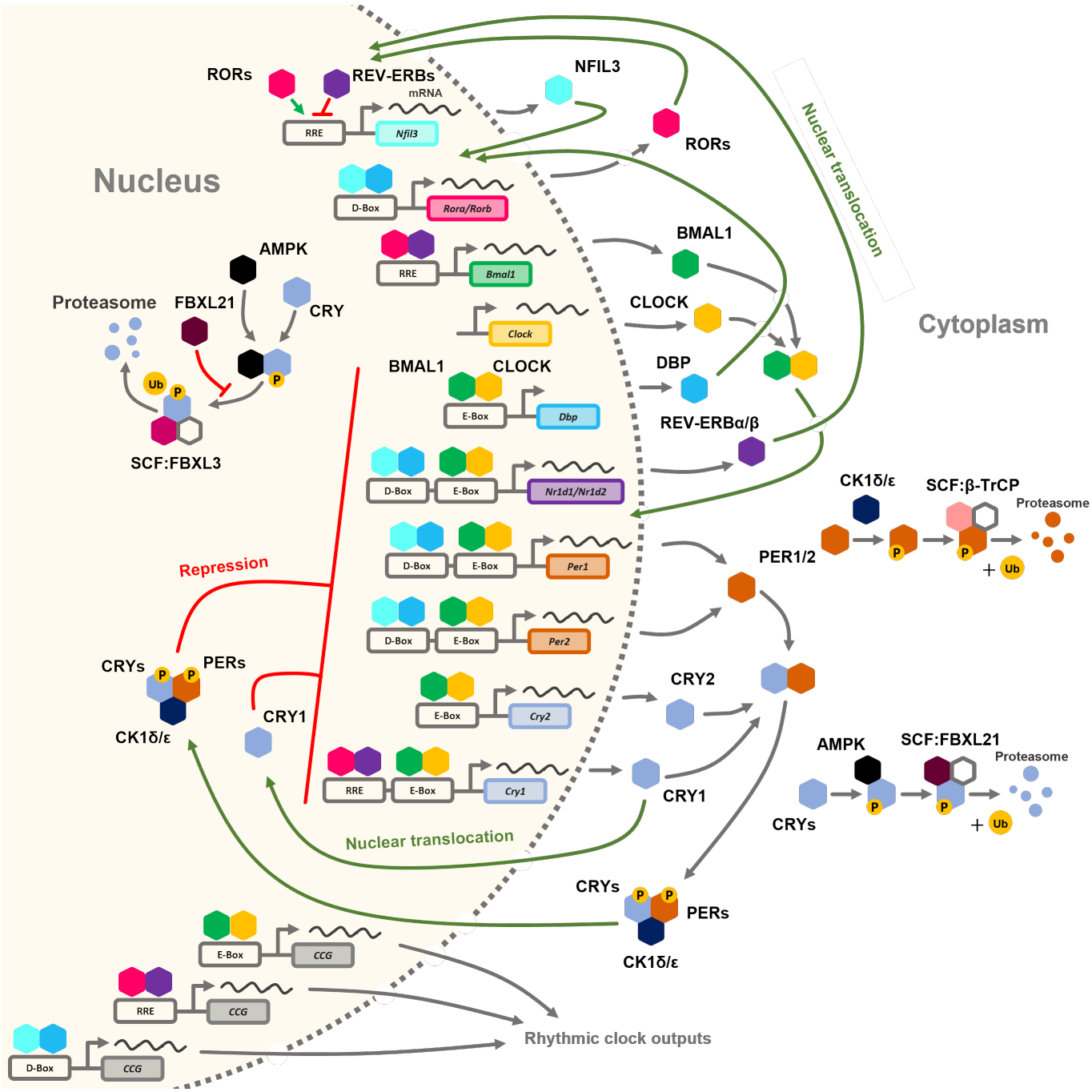

Figure 1.3: The architecture of the mammalian circadian clock operates as a transcription-translation feedback loop. At the core of the circadian network is the heterodimeric transcription factor made up from the proteins CLOCK and BMAL1 which bind and activate E-Box DNA elements. Importantly, these E-Boxes are present in the genes of *Per1*, *Per2*, *Cry1*, and *Cry2*, which are transcribed and translated into proteins which repress CLOCK:BMAL1. This PER and CRY mediated repression is key to the generation of circadian rhythms and occurs via trimerisation with CK1δ/ϵ in the cytoplasm followed by nuclear translocation, and ultimate removal of CLOCK:BMAL1 from E-Boxes by phosphorylation. Recent studies have demonstrated that CRY1 also directly represses CLOCK:BMAL1 without PER proteins. Additional modulating feedback loops are encoded by D-box DNA sites, bound by DBP and NFIL3, and RORE sites, regulated by RORs and REV-ERBα/β, which combine to confer rhythmicity upon BMAL1 as well as downstream clock controlled genes (CCGs). Degradation of PERs and CRYs alleviates repression to begin anew the cycle of transcription and translation. SCF:β-TrCP mediates 26S proteasomal degradation of PER via ubiquitination (Ub) following phosphorylation (P) by CK1δ/ϵ. Whereas, the F-Box binding proteins FBXL21 and FBXL3 antagonistically regulate the degradation of CRY proteins, with FBXL3 importantly promoting degradation of CRYs in the nucleus. Reproduced with permission from Springer Nature and adapted from Takahashi (2017).

regulatory processes (Narumi et al., 2016). One possible answer to this apparent contradiction is that the antibodies might preferentially bind phosphorylated versions of their protein targets and as such protein levels as measured by antibodies are a combined function of both protein abundance and

phosphorylation status. The main source of delay and related pacesetting of the clock is mediated by Casein Kinases 1δ/ϵ (CK1δ/ϵ, encoded by *Csnk1d/e*). PER proteins are bound by CK1δ/ϵ which in turn phosphorylates PERs, marking them for degradation by further promoting polyubiquitination by the Skp1-Cul1-F-box protein and β-transducin repeat-containing protein (SCF:β-TrCP) E3 ubiquitin ligase complex and ultimate degradation by the 26S proteasome (Eide et al., 2005; Shirogane et al., 2005; Reischl et al., 2007), see Figure 1.3. This degradation of PER2 is temperature compensated by a so-called 'phosphoswitch' whereby higher temperatures activates a slow degradation pathway to compensate for generally quicker biochemical reactions (Zhou et al., 2015). Additionally, the ability of PERs to repress transcriptional activity of CLOCK:BMAL1 is also dependent on CK1δ/ϵ (Makoto et al., 2002; Cao et al., 2021). Crucially, nuclear localisation is time-dependent with CRY1 and CRY2 necessary for nuclear entry of PER and CK1ϵ, without which repression by multimeric PER:CRY:CK1 complexes cannot take place, and circadian oscillations cease when both CRYs are absent (van der Horst et al., 1999; Lee et al., 2001; Smyllie et al., 2022). Nuclear localisation of PER1 has also been shown to be inhibited by CK1δ/ϵ dependent phosphorylation of a proximal nuclear localisation signal (NLS) on PER1 close to the CK1 binding site, altering protein confirmation and masking the NLS (Erica et al., 2000). Like PER, the degradation of CRYs ultimately occurs via the 26S proteasome and is also regulated by phosphorylation. Cyclic nuclear entry of nutrient-responsive AMP-activated protein kinase (AMPK) phosphorylates CRY prompting ubiquitination by the SCF:FBXL3 (F-box/LRR-repeat protein 3) E3 complex, further linking metabolism to the clock (Lamia et al., 2009). FBXL3 provides the specificity for degradation as CRY1-2 have been shown to be its substrates, with knock-down of FBXL3 leading to accumulation of CRYs and removal of daily oscillations (Gatfield and Schibler, 2007). In summary, phosphorylation processes modulate protein accumulation, localisation, and function and as such are ultimately responsible for alleviating repression of CLOCK:BMAL1 by PERs and CRYs, beginning the circadian cycle anew. As previously described, mutations to the CK1, such as FASPS or the gain of function tau mutation in the CK1ϵ isoform, drastically alter clock speed, from around 24 hours to 20 hours as the rate of phosphorylation and therefore degradation are increased (Ralph and Menaker, 1988; Lowrey et al., 2000; Vanselow et al., 2006).

Repression by PERIODs and CRYPTOCHROMEs happens in 2 stages

In the nucleus, the principal negative feedback of the core circadian TTFL is repression of transcription activating CLOCK:BMAL1 heterodimers by PER1/2 and CRY1/2 which proceeds via two distinct mechanisms to create three distinct phases of transcription; active, early repressive, and late repressive phases of transcription (Gustafson and Partch, 2015). Evidence shows that PER3 deficient mice do not demonstrate substantial changes in locomotor activity or transcriptional output of key TTFL genes, hence PER3 is not functionally redundant with PER1/2 and has little role in the core TTFL as a transcriptional repressor (Shearman et al., 2000; Spiller et al., 2010; Bae et al., 2001). AT CT8 the active phase of CLOCK:BMAL1 transcription proceeds as described previously by recruitment of transcriptional co-factors CBP and p300, upon alleviation of previous late phase repression by CRY1 through its degradation. Active transcription ends at CT12 with binding by PER:CRY:CK1δ/ϵ complexes which repress CLOCK:BMAL1 by directly lifting it from DNA and inhibiting binding (Koike et al., 2012; Cao et al., 2021). Finally, from around CT0 to CT8, CRY1 initiates the late repressive phase by binding simultaneously to CLOCK PAS-B domain and the c-terminal transactivation domain (TAD) of BMAL1 to block CBP and p300 recruitment to CLOCK:BMAL1 creating a 'poised' state of DNA bound, but transcriptionally inactive, CLOCK:BMAL1 (Ye et al., 2011; Xu et al., 2015). The end of the TAD in BMAL1 appears to be highly conserved amongst vertebrates and insects with vertebrate-like clocks, suggesting this repressive action by CRY is crucial (Yuan et al., 2007). Emerging evidence of differential repressive actions of PERs and CRYs fits well with studies showing spatiotemporal separation of these two repressors.

Ancillary loops

Further feedback loops (Figure 1.3) encompassing Rev-Erbα-β (genes *Nr1d1* and *Nr1d2*), retinoid-related orphan receptors α-γ (RORα-γ), D-box binding protein (DBP), hepatic leukaemia factor (HLF), thyrotroph embryonic factor (TEF), and nuclear factor interleukin-3 regulated (NFIL3, also known as E4bp4) tune the core circadian TTFL and facilitate cross-talk with other cellular systems. Whilst cyclic BMAL1 is not strictly required for generating circadian rhythms it does improve robustness and coupling with output genes (Liu et al., 2008). In brief, CLOCK:BMAL1 drive rhythmic expression of the repressors Rev-Erbα/β which compete with activating RORs at retinoic acid-related orphan receptor (ROR) binding elements in the promoter of Bmal1 (Sato et al., 2004) to yield cyclic BMAL1 protein in antiphase with PER expression (Preitner et al., 2002). Rev-Erb couples BMAL1 to the core TTFL as it is itself activated and repressed by CLOCK:BMAL1 and CRYs/PERs respectively (Preitner et al., 2002). Completing the ancillary loops is DBP, TEF, HLF activating D-box elements, competing with the repressor NFIL3, which has been shown to regulate transcriptional activity of PER1 (Mitsui et al., 2001). These extra loops expand the temporal control that the clock can exert thorough varied phasing via supplemental protein-protein interactions and RORE and D-Box cis-regulatory elements in genes, with regulation by E-boxes central to circadian control of transcriptional (Ueda et al., 2005). Experimentally, these loops have been implicated in coupling with other signalling systems such as immunity (Gibbs et al., 2012) and metabolism (Zhang et al., 2015).

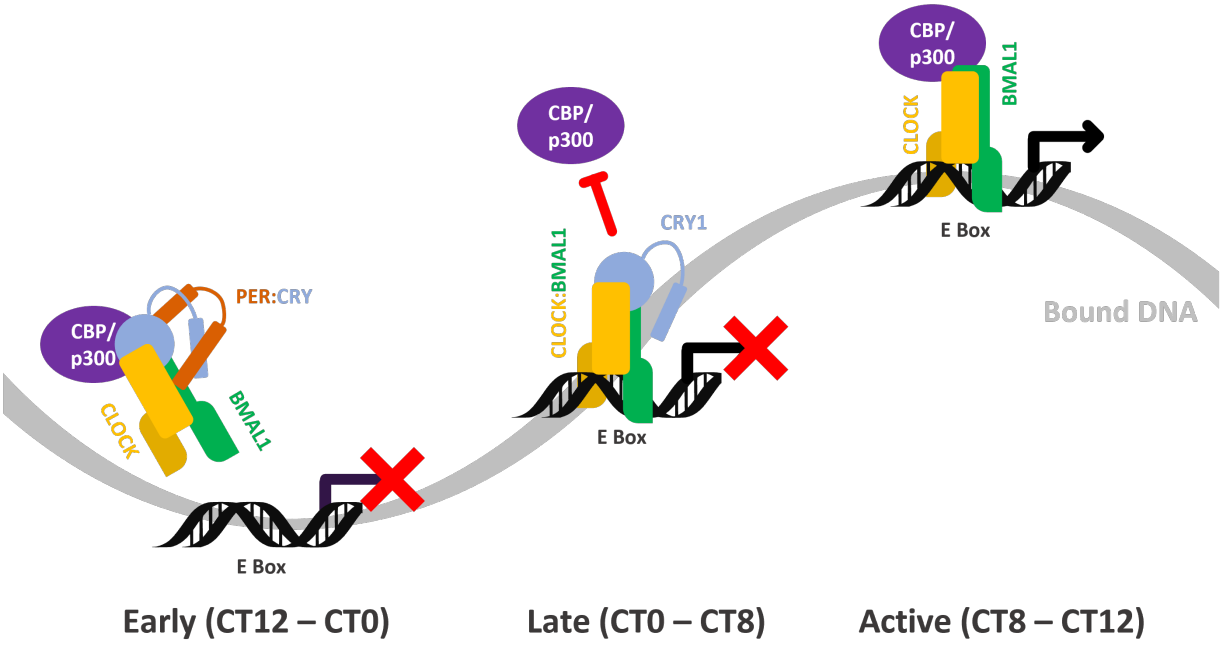

Figure 1.4: Transcription of clock controlled genes by CLOCK:BMAL1 complexes proceeds in 3 distinct phases; active, early repressive, and late repressive phases. PER:CRY complexes act to remove CLOCK:BMAL1 from DNA whereas CRY1 acts alone during the late repressive phase to block transcriptional co-factor CBP and p300 from associating with CLOCK:BMAL1 thus preventing transcription and placing CLOCK:BMAL1 in a 'poised' state. During the active phase CLOCK:BMAL1 can activate transcription of E-Box sites as the repressors have been removed by degradation. Circadian time (CT) refers to the subjective time of the clock, with CT0 defined as the start of activity (lights on for diurnal animals).

1.3 Mathematical modelling

1.3.1 Overview

Mathematical modelling approaches have been used in many previous studies to investigate molecular networks in numerous cell-signalling circuits, spanning inflammatory responses, DNA damage, hypoxic-response, cell cycle, apoptosis, and neuronal differentiation amongst others (Goldbeter, 1991; Fussenegger et al., 2000; Nelson et al., 2004; Bagnall et al., 2014; Chong et al., 2015; Phillips et al., 2016). Depending on what is desired from a model, there are different approaches that may be taken, spanning from models of single molecule interactions (Forger and Peskin, 2003; Gonze and Ruoff, 2021) to whole organisms. Recent studies have used light intensity, heart rate, activity, and temperature measurements taken from personal monitoring devices to predict an individual's circadian rhythms and phase (Huang et al., 2021; Lévi et al., 2020). Each of these systems may be modelled in a variety of ways, such as by solving systems of differential equations, stochastic modelling, and simple oscillatory models (Strogatz, 2018; Forger, 2017). The cellular circadian circuitry has been mathematically modelled in several studies across the last decade in many different organisms. These studies have used both deterministic and stochastic approaches, with notable models by Gonze and Goldbeter (2006) and Forger and Peskin (2003). In this thesis, I build upon the ideas from these models, focusing on how molecules move and interact to generate circadian rhythms as part of the cell autonomous clock mechanism, primarily within fibroblast cells.

To build effective mathematical models, it is essential to appropriately apply statistics, as statistical techniques are used to fit models, decide which models to use, and infer the parameters of a model. One of the main uses of statistics within this thesis is for estimating the parameters of a model based on observed data. I also utilise statistics to discriminate which model out of a set of candidate models best explain the data.

1.3.2 Modelling molecular processes in biology

Mathematically modelling biomolecular processes, such as gene regulatory networks, can aid in understanding the topology (structure) of interactions, infer unseen species, and predict outcomes. It is now understood that regulatory networks are made up of collections of many different motifs that repeat throughout nature, such as simple feed-forward and negative feedback loops (Alon, 2007). At the lowest level, biomolecules can be modelled as particles moving through time and space, which may interact with other molecules depending on their proximity and their electrostatic affinity. The binding of proteins has also been modelled by considering the ensemble of electrostatic forces originating from their constituent amino acids. However, this level of detail is prohibitively difficult to model and simulate for all but the simplest cases. As such, abstractions of these processes are required to fully capture the immense number of molecules present within a cell and the interactions between them. Typically, rates of reaction and collections of molecules are used to simplify models by aggregating individual processes into an average. In this manner, spatial dynamics can be ignored and reactions can be modelled purely in time.

Molecular processes are most often modelled as either mass action kinetics or abstractions of multiple steps such as via Micaelis-Menten or Hill dynamics (Figure 1.5), which define rates of reaction that can be used to construct systems of coupled ordinary differential equations. In the simple case of the law of mass action, the kinetics of a reaction producing species Z from species X and Y binding one another is proportional to the product of the concentration of X and Y, such that

$$\frac{dZ}{dt} = kXY \tag{1.1}$$

where k is the rate constant (Voit et al., 2015). This is often used to represent dimerization reactions. Reactions may also be enhanced by catalysts such as enzymes, which speed up the process of

producing Z from Y, with the rate of reaction dependent on the concentration of X (X is not used up during this reaction). The rate of this reaction was derived for enzymatic activity by Michaelis and Menten in 1913 (Briggs and Haldane, 1925; Johnson and Goody, 2011) and has since been known as Michealis-Menten dynamics. Michaelis-Menten dynamics are a special case of the Hill function whereby the reaction is further mediated by cooperativity, n, between molecules of X such that activation typically only proceeds when n molecules of X act at once. The Hill function is given by

$$Z = \frac{X^n}{K^n + X^n} Y \tag{1.2}$$

for a Michaelis constant K defining the concentration at which X rapidly promotes the reaction, see Figure 1.5. Michaelis-Menten dynamics are recovered from the Hill function when $n = 1$.

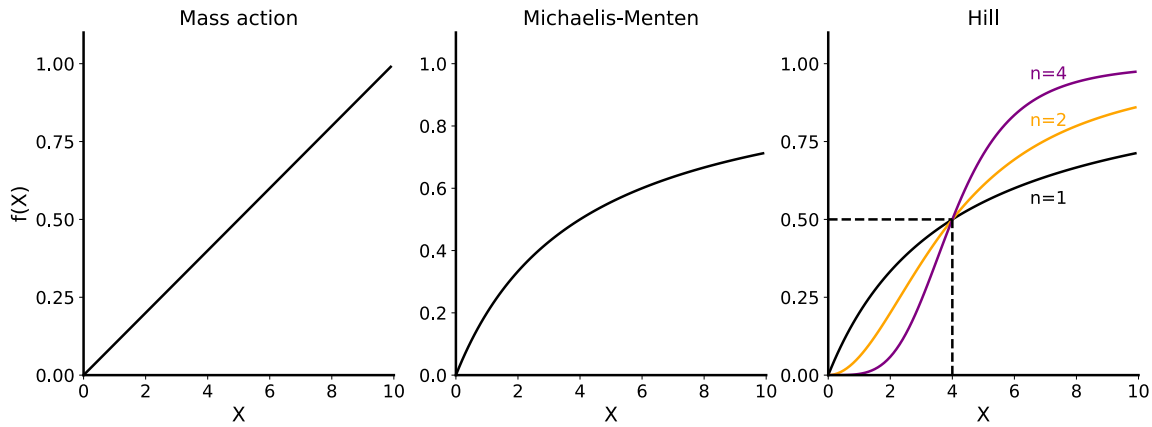

Figure 1.5: Models of reaction rate kinetics. Different rates of reaction over concentrations of X. Mass action is linear in X, whereas non-linear Michaelis-Menten and Hill functions saturate for large values of X compared with the rate constant K. Hill functions become more 'switch-like' as the cooperativity index n increases as seen in the black ($n = 1$), orange ($n = 2$), and purple ($n = 4$) lines. 50% activation is achieved when $X = K$ and can be seen by the black dashed line.

Explicit time delays may also be added to models to represent processes such as expression of a gene following transcription, or a delay in repression after production (Novák and Tyson, 2008). However, pure delays often lead to artefacts and over prediction of phenomena like oscillations as well as being difficult to numerically solve. Recent progress has been made to address these issues using delay adapted Kalman filters (Calderazzo et al., 2019; Burton et al., 2021).

If molecular species abundance is low, as is the case for the usual 2 copies of a genes or a low expression transcription factor, then stochasticity prevails. Commonly used systems of deterministic ODEs cannot capture these heterogenous processes, instead reactions can be translated into sets of reaction processes described by propensities and simulated using a stochastic simulation algorithm such as the Gillespie algorithm (Gillespie, 1976). Alternatively, the hybrid approach provided by the chemical Langevin equations can model stochastic processes as they add a noise terms to ordinary differential equations (Gillespie, 2000).

1.3.3 Parameter estimation and model fitting

When model parameters are not known we may use data to constrain them. This model fitting can be used to find unknown parameters in order to create a predictive model or to infer parameters as an end goal, i.e. finding diffusion rates from confocal microscopy fluorescence recovery after bleaching (FRAP) experiments (Sprague and McNally, 2005). Model fitting, also known as parameter inference, is mathematically defined as an 'inverse problem' where the objective is to determine the parameters of a model that produced the data that we observe. The 'forward problem' of making predictions from a model for a specific set of parameters can often be done quickly either analytically or numerically

solved on a modern computer, however the inverse problem is typically non-trivial for all but the simplest models. Many combinations of parameters often give rise to the same predictions.

In the simplest instance, linear models can be fit with well-defined least squares minimisation algorithms which find the optimal gradient and intercept along with an estimation of the variance on both (Miller, 2006). When models become non-linear they can be fit in a similar manner, but this procedure is especially prone to getting trapped in local minima, overfitting to noise, and underestimation of errors (Motulsky and Ransnas, 1987), especially for high dimensional models with many unknown parameters. Algorithms like simulated annealing and genetic algorithms can overcome issues of fitting to local minima by randomly generating initial parameter guesses (Bertsimas and Tsitsiklis, 1993; Mirjalili, 2019).

Both Frequentist and Bayesian approaches can overcome some of these challenges by implementing a more principled mathematical framework for parameter inference. Statistics provides a set of tools to that can be used to find the most likely parameter values given data or the distribution of potential values in the case of frequentist and Bayesian paradigms respectively (Wood, 2015).

Bayesian inference

The Bayesian approach treats the parameters θ as random variables that follow a probability distribution (Wood, 2015). Initial beliefs about these parameters are also encoded in the 'prior', $p(\theta)$, which when combined and normalised by the 'evidence', $p(y)$, according to Bayes' theorem gives the 'posterior'

$$p(\theta \mid \mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})}, \tag{1.3}$$

where $p(\mathbf{y}|\theta)$ is known as the likelihood. In simple terms, this is the probability over the parameters given the data and our prior beliefs. The likelihood can be thought of as the model of the system, taking in parameters θ and yielding a distribution of possible data y. Additionally, the evidence is given by

$$p(\mathbf{y}) = \int_{\theta} p(\mathbf{y} \mid \theta)p(\theta)d\theta \tag{1.4}$$

and in many cases it is hard or even impossible to integrate this term. Fortunately, as it only normalises the posterior, the marginal likelihood is commonly ignored. Defining the likelihood, $p(\mathbf{y}|\theta)$, is often the most difficult task when utilising this framework as the model may not be analytically tractable as it often defined by an integral. When this typically high dimensional integral cannot be solved in closed form it may instead be sampled via Markov chain Monte Carlo techniques such as Metropolis-Hastings random walks (Hastings, 1970) or Hamiltonian Monte Carlo (HMC) (Betancourt, 2017). In MCMC, a Markov chain is constructed that explores the parameter space and generates a set of samples from the posterior distribution. The samples can then be used to make inferences about the parameters, such as the mean and variance of the posterior distribution. MCMC is guaranteed to provide the solution given enough iterations, however it is non-trivial to setup and is computationally intensive (Betancourt, 2017).

Inferring parameters in this fashion has gained popularity in recent years for its principled mathematical underpinnings and promise of truthful uncertainty quantification given noisy data. Bayesian inference has been applied in many areas and - closer to the aims of this thesis - for inferring parameters governing transcription of mRNA (Jenkins et al., 2013; Hey et al., 2015; Featherstone et al., 2016; Gómez-Schiavon et al., 2017).

Frequentist approach: maximum likelihood estimation

In the frequentist approach, parameters are estimated based on the maximum likelihood principle such that parameter values that make the observed data more probable are more likely to be true. Hence the name *maximum likelihood estimation* (MLE). Similarly to Bayesian inference, in the frequentist framework, probability distributions, are also used to the model the chance of observing the data, $\mathbf{y} \in Y$, given the parameters, $\theta \in \Theta$, where Y and Θ are Banach spaces (Stuart, 2010). This probability is often written as $p(\mathbf{y}|\theta)$ (Wood, 2015). As the data is a known quantity that is fixed, the probability distribution of θ given $\mathbf{y}$, known as the *likelihood* $L(\theta|\mathbf{y})$, is equal to the model, such that

$$L(\theta|\mathbf{y}) = p(\mathbf{y}|\theta). \tag{1.5}$$

When considering complex models, $p(\mathbf{y}|\theta)$ may be made up from joint distributions, thus it is common to work with the log-likelihood, $l(\theta|\mathbf{y}) = \log L(\theta|\mathbf{y})$, to convert multiplicative terms into summations of logarithms of distributions. Finally, the point estimate for the parameters θ is found by maximising this log-likelihood

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}}\ l(\theta|\mathbf{y}). \tag{1.6}$$

In practice this procedure is often accomplished via computational methods. However, if the likelihood is differentiable over the space Θ, then the maxima of l with respect to θ can be determined by differentiating as

$$\frac{\partial l}{\partial \theta}\Big|_{\hat{\theta}} = 0 \tag{1.7}$$

(Haynes, 2013). MLE is generally equivalent to the maximum a posteriori (MAP) estimation with uniform prior distributions when viewed within a Bayesian inference framework. In general, MLE is asymptotically normal and is an efficient estimator as the sample size tends to infinity (Pfanzagl, 1994).

Choosing the right model with model selection

When constructing a predictive mathematical model of a process, there are often a number of possible candidate models that might explain the data. Some models can describe the data more effectively by employing a larger number of parameters to capture the underlying process more precisely. However, highly complex models tend to face identifiability issues; the true value of certain parameters may remain elusive, even with infinite observations. One solution is to increase the dimensionality of the data, enabling the model to fit to more of the outputs of the process being modelled. If that approach is unfeasible, a simpler model that can adequately fit the available data becomes preferable. Identifying the most parsimonious model presents a non-trivial problem, and it often falls to the modeller's intuition to strike an appropriate balance between model complexity in the context of available data (Forger, 2017). Various statistical techniques exist to find the most optimal model through model selection. One notable measure is the Akaike Information Criterion (AIC), which assigns higher scores to models based on their data fit, adjusted for the number of free parameters. The AIC score is defined as

$$\mathrm{AIC} = 2k - 2\ln \hat{L} \tag{1.8}$$

where k is the number of free parameters and $\hat{L}$ is the maximum likelihood (Akaike, 1974). In the case of non-linear least squares fitting where errors are distributed according to independent and identical normal distributions, $\hat{L}$ is equal to the sum of squared errors. Lower AIC scores indicate the candidate model is more suitable as models with too many parameters are penalised. A similar

measure called the Bayesian Information Criterion (BIC) was developed shortly after by Schwarz (1978) and is a harsher measure on the number of free parameters by increasing the penalty with the number of data points n from $2k$ to $k \log(n)$, where n is the number of data points. As such BIC reduces the likelihood of overfitting to data, a common problem in high dimensional models with lots of data. The BIC is similar to the AIC, but places a stronger emphasis on model complexity. The BIC rewards models that fit the data well and heavily penalises models with a large number of parameters. The advantage of BIC is that it is less biased towards models with many parameters than the AIC. However, it can be overly stringent and may not be suitable for all types of data. Both the AIC and BIC are based on the maximum likelihood principle and assume that the likelihood provides a valid approximation. Both AIC and BIC are asymptotically unbiased and perform well for large sample sizes. For small sample sizes, the corrected AIC score, AICc, with an extra penalty for the number of parameters, is used (Hurvich and Tsai, 1989). AIC consistently selects models with better predictive power. However, it is not guaranteed to choose the true model from a set of candidate models, even with an infinite number of observations. In contrast, BIC excels at selecting the true model, as it is guaranteed to do so given enough data (Chakrabarti and Ghosh, 2011). In either case, the uncertainty in the estimates of the parameters is not accounted for. In summary, solely relying on these criteria can lead to the selection of an inappropriate model. It's essential to choose models that not only are useful but also capture uncertainty accurately.

Cross-validation can also be used to assess the ability of a model to generalise to new data. Data are divided into a training and validation sets, with the model fit to the training set and its predictions evaluated on the validation set. The advantage of cross-validation is that it provides an estimate of model performance on new data, but it can be computationally intensive for complex models and requires a large sample size. Obtaining a large enough sample size for cross-validation from biological experiments is often difficult, limiting its use. All of these techniques have their advantages and disadvantages when used in practice. It has been shown that leave-one-out cross-validation is asymptotically equivalent to AIC (Stone, 1977). BIC is also asymptotically equivalent to leave-k-out cross-validation (Shao, 1997).

1.3.4 Current mathematical models of the clock

Analysing oscillatory timeseries data

How can we use oscillatory models to deepen understanding and answer research questions? In the study of biological rhythms, we are often concerned with whether timeseries data, such as from ChiP-seq (Koike et al., 2012) or bioluminescent tagged proteins (Yoo et al., 2004), are cycling and if so with what period, amplitude, and phase. Several tests and computer algorithms have been developed for this purpose, such as the statistical models based on based on spectrum resampling (Costa et al., 2013), Fisher's G statistical test (Fisher, 1929; Wichert et al., 2004) and JTK_CYCLE (Hughes et al., 2010) as well as those fitting cosine functions such as COSOPT (Panda et al., 2002a) and ARSER (Yang and Su, 2010). These types of analyses are useful in exploratory analyses of putative oscillatory data, as they do not make assumptions about the underlying oscillatory structure arising from molecular or cellular interactions. Insights gained from this can be used to inform hypotheses, further experimentation, and mechanistic models.

Biological oscillators and limit cycles

Lotka and Volterra studied oscillations in biological and chemical systems by deriving differential equations with undamped solutions from examining the interactions between a predator and prey species as well as the law of mass action respectively (Lotka, 1920). In the 1960s Brian Goodwin began studying biological oscillators in mathematical detail following reports of feedback inhibition of genes (Jacob and Monod, 1961). Goodwin constructed an ODE model of two variables, X and Y,

which represented mRNA, X, and its protein product, Y, evolving through time, see Figure 1.6(A-C), in which Y represses the production of X according to the non-linear term $K/(K+Y)$, for constant K (Goodwin, 1965). Numerical integration of these equations, such as via the Runge-Kutta method, yields oscillations, see Figure 1.6D. The non-linearity of repression is vitally important as the level of repression is initially small for small Y, rapidly increasing when $Y \approx K$, and then saturating, without which there is insufficient delay in the system to sustain oscillations (Novák and Tyson, 2008). Goodwin went on to model how two oscillators may couple and interact, finding that, depending on specific choices of parameters, the two may entrain one another, synchronise in different phases or quench the other, foreshadowing many experimental findings to come from coupled biological oscillators such as cells in the SCN.

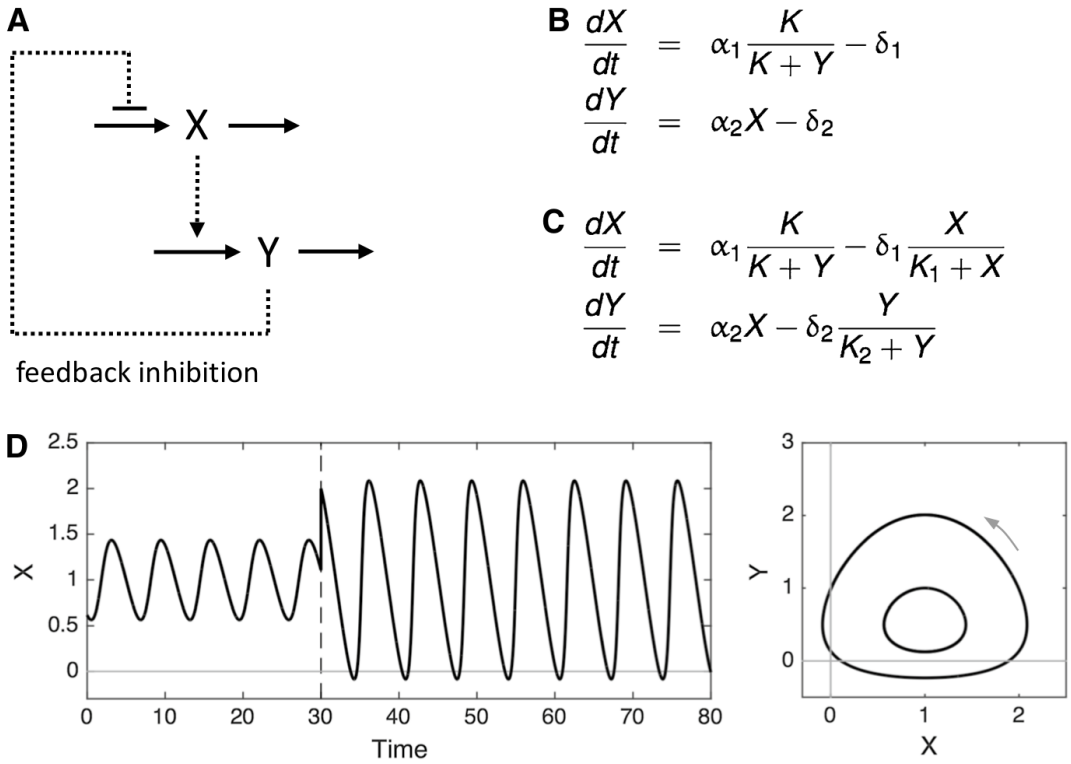

Figure 1.6: The two-variable Goodwin model. (**A**) Model of an mRNA, X, translated into protein, Y, which represses the production of X. (**B**) Coupled ordinary differential equations defining the model in (A) with linear degradation terms. (**C**) A modified form of the model in (**B**) whereby degradation follows Michaelis-Menten dynamics. (**D**) Oscillations obtained by numerical integration of the two-variable model given in (**B**), for the following parameter values: $\alpha_1 = 2, \alpha_2 = 1, \delta_1 = \delta_2 = 1, K = 0.5$. At time t = 30 the value of the variable X was increased. Reproduced with permission from Springer Nature and adapted from Gonze and Ruoff (2021).

Stable biological oscillators, like the circadian clock, should be insensitive to small perturbations and not overly reliant on the initial conditions of the system. Goodwin's original two state oscillator does not fulfil these conditions, see Figure 1.6, as any perturbation will permanently shift the oscillator into a new regime, hence Goodwin sought to correct this problem. Goodwin solved these issues by adding in a third component, Z, which conferred the system self-sustaining oscillations that are protected against perturbations. This type of oscillatory regime is known as a stable limit cycle, whereby a stable orbit exists that trajectories close to will decay to (Strogatz, 2018). Stable limit cycles are now widely accepted to describe many biological oscillators (Goldbeter, 1996; Forger, 2017) and have been used to explain counterintuitive phenomena. Stochastic phasing of the circadian clock between individual *Cyanobacteria* after temperature pulses has been explained by the proximity of the trajectory to the singularity after perturbation, leading to either close or random phasing of individual oscillators depending on the phase when the pulse was given (Gan and O'Shea, 2017).

Non-linearity, cooperativity, and multiple loops are required for stable oscillations

Part of Goodwin's successful creation of a stable limit cycle oscillator with 3 variables was the use of a Hill function ($f = K^n/(K^n + Z^n)$), which modified the Michaelis-Menten type dynamics (when $n = 1$) by the degree of cooperativity (defined by n) required to repress production of X, see Figure 1.5. Cooperativity is the simultaneous number of components required to initiate or repress a process and in the 3 component Goodwin oscillator it has been shown that an implausibly high degree of cooperativity ($n > 8$) is required for the limit cycle to exist (Griffith, 1968). Fortunately, modification and addition of non-linear terms in degradation, production, and repression all aid in reducing required cooperativity. In the case of gene regulatory networks, cooperativity represents multiple possible processes, examples include multiple binding sites within a genes promoter, formation of complexes to either activate or repress the gene (Keller, 1995), and multi-site phosphorylation (Gunawardena, 2005). It is perhaps no surprise then that we find significant use of phosphorylation in cellular control mechanisms in addition to dimerization in many biological systems from NF-κB binding its repressor IκB to CLOCK:BMAL1 binding with PERs and CRYs.

In practice it is rare to find a bio-molecular process that represents a Hill exponent higher than 4, hence additional components and loops are often included to widen the parameters over which an oscillator will function. Adding in additional steps that are required before repression, improves the likelihood of oscillations by reducing the minimum required cooperativity. The greatest reduction in the required cooperativity is seen when two or more oscillators are coupled together (Kurosawa et al., 2002). A study comparing saturation of 'in-loop' kinetics and addition of new processes to the loop such as degradation or nuclear import/export – so called branching processes – found branch reaction steps best promotes oscillations (Kurosawa and Iwasa, 2002). This has been further observed in countless other studies where additional coupled negative or positive loops aid stability and generation of rhythms as well as conferring specific properties. Negative feedback loops have been shown to have higher sensitivity to amplitude perturbations whilst exhibiting lower period sensitivity than positive feedback loops (Baum et al., 2016). Hence, stable biological oscillators are usually delayed negative feedback loops as it is hard to perturb their period, whereas rhythms in the heart and cell cycle instead utilise feedforward loops to tune their frequency whilst maintaining consistent output amplitude (Tsai et al., 2008). Combinations of positive and negative feedback loops can expand the range of parameters over which oscillations are permitted, explaining the combination of negative feedback (PERs and CRYs) and positive feedforward loops (Rev-Erbs and RORs) in the circadian clock (Relógio et al., 2011). Overall, these studies demonstrate how long period circadian oscillations – much longer than the half-life of the most long-lived oscillatory component – are built up from multiple fast processes occurring on the order of seconds to minutes.

Are circadian clocks damped noisy oscillators?

All molecular processes demonstrate noise, also known as stochasticity, resulting from both intrinsic and extrinsic sources. Biological oscillators are no exception, which under normal operation are stable against small perturbations as previously discussed. However, when molecular numbers are low, random noise can become comparable to the mean abundance or significantly perturb the future trajectory of the system. Gonze and Goldbeter demonstrated via stochastic simulations that as the system size (or total molecular abundance) of a simple 3 component oscillator reduced, the greater the relative noise became. Thus increase in noise also increased the variance of the period and eventually removed rhythms altogether (Gonze and Goldbeter, 2006). Interestingly, noise can also play a part in the emergence of self-sustained oscillations as stochastic transcription, translation, and degradation may permit Hill exponents of 1 (Tiwari and Fraser, 1973). The limit of 2 gene copies per cell and measurements of low molecular abundance for the clock (Narumi et al., 2016; Smyllie et al., 2016) suggests that stochasticity must indeed play a role in the circadian clock (Li et al., 2020).

Damped oscillations are often observed in dispersed cells in culture, and in decoupled SCN neurons

it has been shown to improve entrainment to a wide range of periods (Webb et al., 2012). It has been theoretically argued that a damped oscillator is better able to entrain to external cues than one with a stable limit cycle which can oscillate forever without external input (Woller et al., 2014). This has a clear interpretation for circadian clocks within multicellular eukaryotes (i.e. mammals); weak, coupled, and easily entrainable oscillators possess superior qualities for coordinating 24-hour processes rather than strong and rigid oscillators (Gonze et al., 2005; Komin et al., 2011). A recent study by Unosson et al. (2021) demonstrates that the circadian oscillations of cells in the dorsal region of the SCN shift more upon entrainment when compared with the central region. This finding highlights the balance between robust rhythms and the ability to entrain to environmental cues.

Models of the mammalian clock predict rapid DNA binding and strong protein-protein interactions

Biologically accurate and expansive mathematical models of the mammalian circadian clock have revealed common motifs and design principles upon which these clocks are built (Leloup and Goldbeter, 2003; Mirsky et al., 2009). As previously discussed, combinations of additional loops and coupling improve robustness of the clock and its tunability, which seems essential for proper clock regulation in mammals (Becker-Weimann et al., 2004; Relógio et al., 2011). Forger and Peskin also introduced a deterministic ODE model (Forger and Peskin, 2003) of the mammalian clock which included the core TTFL (PER/CRY), which was followed up by a stochastic version (Forger and Peskin, 2005). Interestingly, for their stochastic model to produce robust 24-hour rhythms, similar to the original deterministic model, required fast kinetics on the promoter in the order of seconds, accurately predicting findings that came years later that showed CLOCK:BMAL1 binding to the DBP promoter of approximately 10 seconds (Stratmann et al., 2012). Moreover, duplications of similar genes with similar functions such as the two isoforms of PERs and CRYs further improved this effect. Forger and Peskin also found that decreasing the number of nuclear transcription factors (TFs) exponentially increased the variance of the period from around 1 hour for 4000 nuclear TFs to approximately 3 for 500 TFs. Kim and Forger (2012) also showed that for the mammalian TTFL the 1-1 stoichiometry between repressors (PERs and CRYs) and activators (BMAL1, CLOCK/NPAS2) is vital to robust circadian rhythms. Outside of a narrow range of stoichiometries the clock is rendered arrhythmic. Strikingly, the tighter the binding, $K_D = 10^0 - 10^3$ nM, between the repressors to the activators the wider the range of allowed stoichiometries. Furthermore, strong affinities between repressive and activating proteins confers ultra-sensitivity to changing levels of repressors, permitting fast switching from repression to activation.

1.3.5 Diffusion: modelling molecular over time and space

All molecules, including proteins and RNA, must diffuse across the cellular environment to perform their necessary functions. This molecular movement can be modelled by its diffusion characteristics, defined by the diffusion coefficient D and anomalous parameter α. The diffusion coefficient links the mean squared displacement (positional variance) with the elapsed time as

$$\text{MSD} = 4\pi D(\Delta t)^{\alpha}. \tag{1.9}$$

Normal diffusion occurs when $\alpha = 1$ (Einstein, 1905; von Smoluchowski, 1906). Anomalous diffusion, characterised by $\alpha \neq 1$, describes both sub- and super-diffusion, the former being due to confinement and molecular interactions with slow or immobile substrates whereas super-diffusion is due to active transport or flow (Höfling and Franosch, 2013), see Figure 1.7A. Thus, quantifying the diffusion of different molecules within living cells can elucidate how these molecules are interacting with the cellular environment.

Contrary to normal diffusion (known as Brownian motion), molecules can interact with the cellular environment in complex ways. This can include interactions with other molecules, barriers, obstacles, and active transportation mechanisms. Within the cytoplasm molecules can be crowded out, confined

to membranes, or actively transported. In the nucleoplasm molecules must contend with packed DNA and the nuclear matrix crowding the space. If a molecule binds DNA then it will interact with the long and winding DNA strands in complex patterns by sliding, hopping, and transferring between strands (Schmidt et al., 2014), see Figure 1.7B.

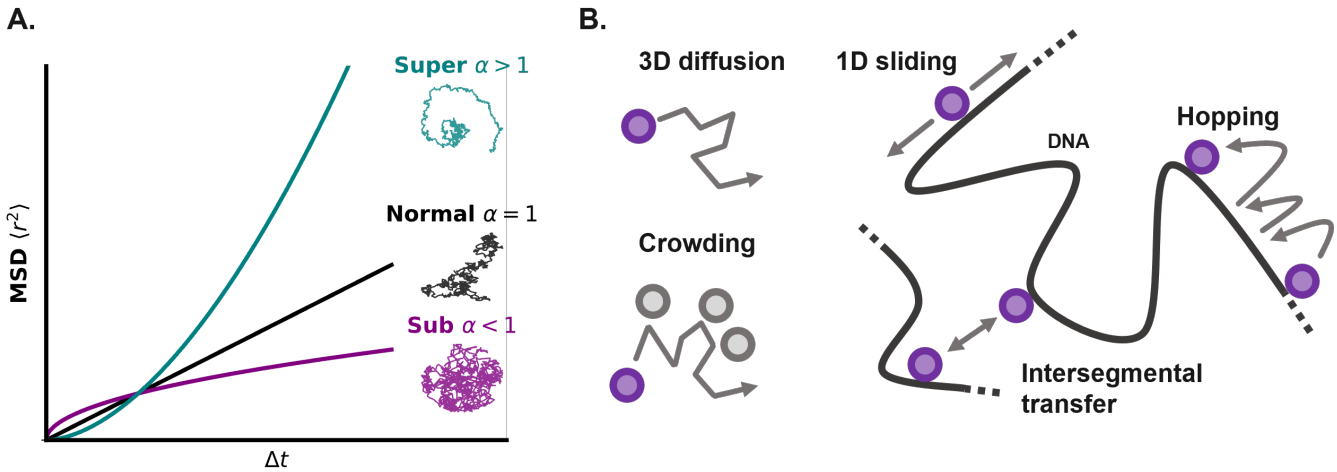

Figure 1.7: Diffusion. (**A**) Mean squared displacement (MSD) from an initial position is characterised by the elapsed time, Δt, diffusion coefficient D, and anomalous exponent α as $\langle r^2 \rangle = 4\pi D(\Delta t)^\alpha$. Inserts show representative molecular trajectories for super-diffusion (teal), normal diffusion (black), and sub-diffusion (purple). (**B**) Modes of DNA binding proteins diffusing in the nucleus, including normal 3D diffusion, molecular crowding in 3D, 1D sliding and hopping, and 2D intersegmental transfer. Adapted from Schmidt et al. (2014).

DNA binding molecules sub-diffuse with anomalous exponents below one, decreasing in relation to their extent of interaction with DNA, see Figure 1.7B. Even without binding DNA, molecules within the nucleus are confined by the crowded conditions. Several modes of movement for DNA binding molecules exist, with the simplest defined by histones which semi-permanently bind DNA, and the most complex by transcription factors (TFs). TFs search the genome for their complimentary binding sites by executing a 'linear search', landing on DNA and sliding along the DNA in 1 dimension. TFs may also hop along strands of DNA or transfer between different strands to further speed up their search for their binding sites. Without these much faster 1D and 2D modes TFs would likely be incapable of properly regulating the full repertoire of genes across the billions of base pairs found in mammalian genome if they only relied on normal 3D diffusion (Schmidt et al., 2014).

1.4 Experimentally quantifying molecular systems

1.4.1 Fluorescence microscopy: observing protein with molecule 'light bulbs'

Live-cell protein reporters give researchers access to longitudinal measurements in bulk or single cells, without the multiple drawbacks of antibody based biochemical protein assays such as western blotting and immunohistochemistry. Both luminescent and fluorescent protein reporters provide dynamic, temporal assays of protein abundance.

The luminescent protein luciferase (LUC) has been extensively used to track protein expression or promoter activity when measured using sensitive photomultiplier tubes (PMT) and provide the most sensitive tag for imaging as the only signal in dark conditions will be due to the luciferase interacting with its substrate luciferin, ATP, and oxygen to produce photons (Spiller et al., 2010). Yet, luciferase imaging is limited by its long integration times ($\sim$10 minutes) reducing its applicability for studying fast processes. Fluorophores are far more useful when investigating protein dynamics due to the superior spatial and temporal resolution in addition to a multitude of colours available for tagging. Furthermore, powerful confocal microscopes are now capable of detecting less abundant tagged

proteins by using very sensitive detectors and overcoming issues of auto-fluorescence by spectral or fluorescent lifetime separation. Fluorescent proteins are also amenable to study through quantitative techniques, such as fluorescence recovery after photobleaching (FRAP), fluorescence correlation spectroscopy (FCS), and fluorescence cross-correlation spectroscopy (FCCS).

The first fluorescent protein, Green Fluorescent Protein (GFP), was isolated from the jellyfish *Aequorea Victoria*, which exhibited green fluorescence (emission $\lambda_{\text{max}} = 509$ nm) when exposed to UV/blue light (Prasher et al., 1992). Shortly after it was shown that GFP could be fused to another protein via genetic manipulation to create stable fluorescent fusion proteins (Chalfie et al., 1994). The discovery and development of GFP as a reporter earned Roger Y. Tsien, Osamu Shimomura, and Martin Chalfie the Nobel prize in Chemistry in 2008. A wide-ranging catalogue of fluorescent proteins now exists for imaging over a wide range of wavelengths from blue to far-red that are codon optimised for expression in mammalian cells (Lambert, 2019).

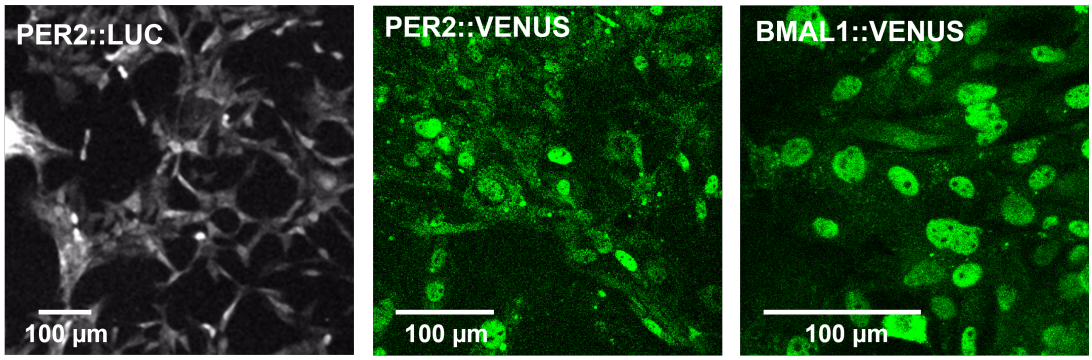

Figure 1.8: Confocal microscopy images of primary fibroblasts in culture that have been genetically engineered to express (**left**) PER2:LUC luminescence, (**middle**) PER2::VENUS fluorescence, and (**right**) BMAL1::VENUS fluorescence.

Luminescence and fluorescence have been utilised to visualise several components of the circadian clock. Circadian experiments often used the fusion reporter PER2::LUC, with oscillations in bioluminescence found in explants of the SCN as well as peripheral tissues, including liver, lung and heart (Yoo et al., 2004; Liu et al., 2007; van der Veen et al., 2012). Imaging of tissue explants over 6-weeks have revealed robust oscillations that don't dampen over time, merely becoming desynchronised within a population (Leise et al., 2012). Recently, quantification by FCS of key proteins PER2 and BMAL1 has been facilitated via fusion with the fluorescent protein Venus (Smyllie et al., 2016; Yang et al., 2020), see Figure 1.8.

1.4.2 Measuring protein binding to DNA by watching its movement using Fluorescence Recovery After Photobleaching

Fluorescence Recovery After Photobleaching (FRAP) is a single-cell bulk measurement of the behaviour of fluorescent molecules (here typically fluorescent fusion proteins) developed in 1976 (Axelrod et al., 1976; Koppel et al., 1976) and has since seen successful used for in vivo measurements (Sprague and McNally, 2005). In FRAP, high intensity laser light is used to permanently bleach fluorescent molecules within a region of interest (ROI) to observe the resulting recovery of fluorescence into the ROI, Figure 1.9A. This recovery reveals kinetic information about the molecules as they diffuse into the ROI from surrounding non-bleached areas. Fluorescence Loss in Photobleaching (FLIP) is a derivative of this technique which examines loss of florescence in a non-bleached portion of the cell, i.e. the cytoplasm, after bleaching a different cellular compartment such as the nucleus, which in this example would yield the nuclear import rate (Spiller et al., 2010).
Exact recovery profiles for FRAP vary due to the size of the bleached ROI and how labelled proteins move, slower diffusing proteins (larger molecules) recover slower than those with high diffusion coefficients. Furthermore, different regimes of molecular movement, such as DNA binding and con-

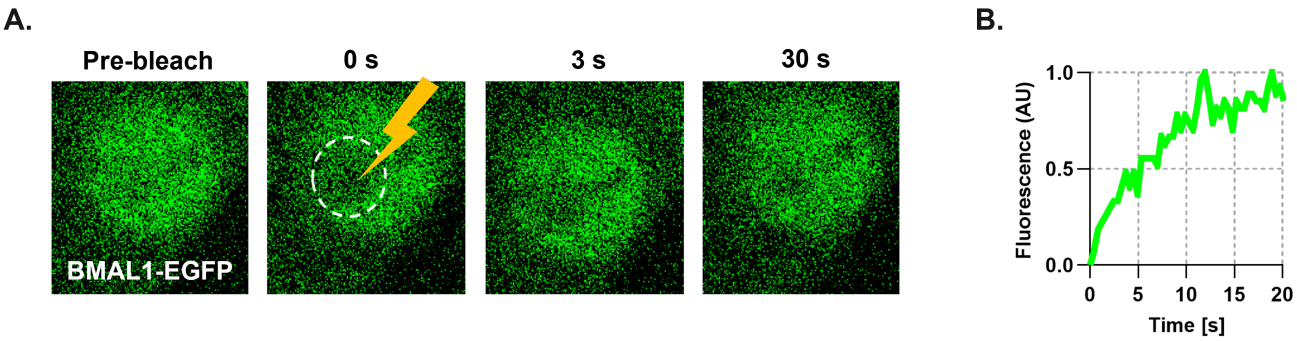

Figure 1.9: Fluorescence recovery after photobleaching. (**A**) Confocal microscopy images of BMAL1:EGFP before and after photobleaching of a region (white dotted circle) (**B**) Representative recovery of fluorescence within the bleached region. The speed of recovery is related to the diffusion of BMAL1:EGFP binding with relatively immobile DNA.

finement, also alter recovery profiles in a distinguishable manner (Sprague and McNally, 2005). For example, free GFP recovers in under a second and is well fit by a simple diffusion model, whereas when fused with certain proteins fluorescence can take seconds or even minutes to recover (Carrero et al., 2003). Processes like DNA binding can uncouple diffusion from the recovery process as the rate of binding is significantly slower than diffusion. This can be understood as non-observable fluorescent fusion proteins that have bleached fluorophores 'blocking' binding sites until they vacate the site according to their characteristic off rate k_{OFF}. Diffusion can often be neglected when fluorescent molecules move across the ROI much faster than the binding kinetics, in which the recovery curve is fit by

$$f(t) = \frac{1 - k_{\text{ON}}}{k_{\text{ON}} + k_{\text{OFF}}} e^{-k_{\text{OFF}} t} \tag{1.10}$$

where k_{ON} is the association rate (Sprague et al., 2004), see Figure 1.9B. In this regime the two rates rates are strictly positive, with the effective on-rate k_{ON} typically between $10^2 - 10^8$ M^{-1} s^{-1}. Additionally, the off-rate k_{OFF} has been measured to be of the order 10^{-4} to 10^0 s^{-1} for chromatin remodelling proteins and transcription factors respectively (Hettich and Gebhardt, 2018). In most analyses of DNA binding by FRAP finding the association rate is neglected and only k_{OFF} is considered.

As FRAP is a bulk measurement of many proteins across a significant proportion of the cell (>10%), it can suffer from identifiability issues when selecting a model. Proteins that exhibit different modes of movement at similar frequencies will be averaged over, obscuring which model is the best fit to the data. Hence, knowledge about expected dynamics and controls involving unfused fluorescent molecules are crucial to reliable quantification of FRAP measurements (Sprague and McNally, 2005).

1.4.3 Fluorescence Correlation Spectroscopy: counting photons from fluorescent proteins to quantify their concentration, diffusion, and interactions

Fluorescence correlation spectroscopy

Fluorescence correlation spectroscopy (FCS) is a powerful microscopy technique that determines the absolute count of fluorescent molecules in a solution or within a cell. This determination is achieved by analyzing the fluctuations in the number of photons observed within a small confocal volume, approximately 1 fL (Schwille et al., 1997) (Figure 1.10A). In the FCS setup, a pinhole blocks out-of-focus light, ensuring that fewer than 1000 molecules contribute to the observable signal at any given moment (Liu et al., 2015). Intuitively, the total rate of observable photons is directly proportional to the number of molecules, N, and the brightness per molecule. This principle is employed by the closely related photon counting histogram (PCH) technique (Chen et al., 1999). However,

accurately measuring molecular brightness presents challenges, primarily because it necessitates calibration against known fluorophore concentrations. Consequently, the applicability of PCH in in vivo studies is restricted as the brightness of a fluorescent molecule can be influenced by its surroundings. For instance, fluorescent proteins display varying emission characteristics across different pH levels (Kneen et al., 1998).

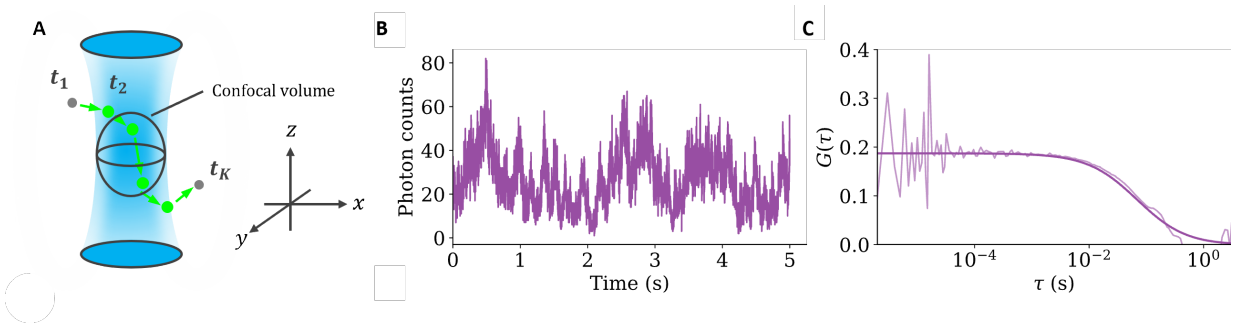

Figure 1.10: Fluorescence correlation spectroscopy. (**A**) Schematic of green fluorescent molecules diffusing in a small illuminated confocal volume (shaded). (**B**) Binned photon counts detected from fluorescent particles diffusing within the illuminated confocal volume. (**C**) Auto-correlation of the counts over progressively longer lag times. an interacting pair of proteins. The dark line represents the model fit to the auto-correlation data.

FCS addresses the challenge of molecular brightness by analyzing the fluctuations in the photon signal over time, as illustrated in Figure 1.10B. It uses these fluctuations to infer the number of molecules and their diffusion rate, eliminating the need to know the brightness of individual molecules. The technique involves auto-correlating the signal across successive lag-times, specifically excluding a lag time of zero to avoid self-correlation. In this auto-correlation process, the number of observable photons shows a high correlation at shorter lag times, diminishing as the lag time between the compared time bins increases, as depicted in Figure 1.10C. This diminishing correlation arises from molecules diffusing in and out of the observation volume. The auto-correlation at short lag times inversely relates to the average number of molecules in the confocal volume. Meanwhile, the declining slope of the auto-correlation provides insights into the diffusion rate, D, of the molecules. Particles that diffuse more slowly remain correlated for extended durations, causing a rightward shift in the curve. If the fluorescent particles diffuse slower then they remain more highly correlated for longer periods of time, thus shifting the curve to the right. The auto-correlation, $G(\tau)$, of the photon intensity, I, at time t as a function of different lag times τ is calculated as

$$G(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle} - 1 \tag{1.11}$$

where angled brackets, $\langle\cdot\rangle$, denote the time average. The lag times are greater than 0, thus ensuring that $0 < G(\tau) < 1$. In the case of simple 3D Brownian motion, the auto-correlation of photons given a lag of τ can be modelled as

$$G(\tau) = \frac{1}{\langle N\rangle}\left(1+\frac{\tau}{\tau_D}\right)^{-1}\left(1+\frac{\tau}{S^2\tau_D}\right)^{-\frac{1}{2}} \tag{1.12}$$

where τ_D is the diffusion dwell time, S defines the shape of the confocal volume such that $S = w_z/w_{xy}$, and $\langle N\rangle$ is the average number of particles in the detection volume (Tian et al., 2011). The radii of the Guassian confocal volume in the horizontal axes, x and y, is symmetrical and given by w_{xy}, with the radius in the vertical, z, direct given by w_z. The diffusion dwell time is related to the diffusion rate as

$$D = \frac{w_{xy}^2}{4\tau_D}. \tag{1.13}$$

Models of the auto-correlation are typically fit to data via least squares χ^2 minimisation, thus giving a point estimate of the parameter values. Beyond simple normal diffusion, there exist a number of different theoretical models of correlation curves that can be used to estimate, anomalous diffusion coefficients (Höfling and Franosch, 2013), chemical relaxation times (Haupts et al., 1998), triplet states (Widengren et al., 1995), and protein-DNA binding kinetics (Michelman-Ribeiro et al., 2009).

Measuring interactions with fluorescence cross-correlation spectroscopy

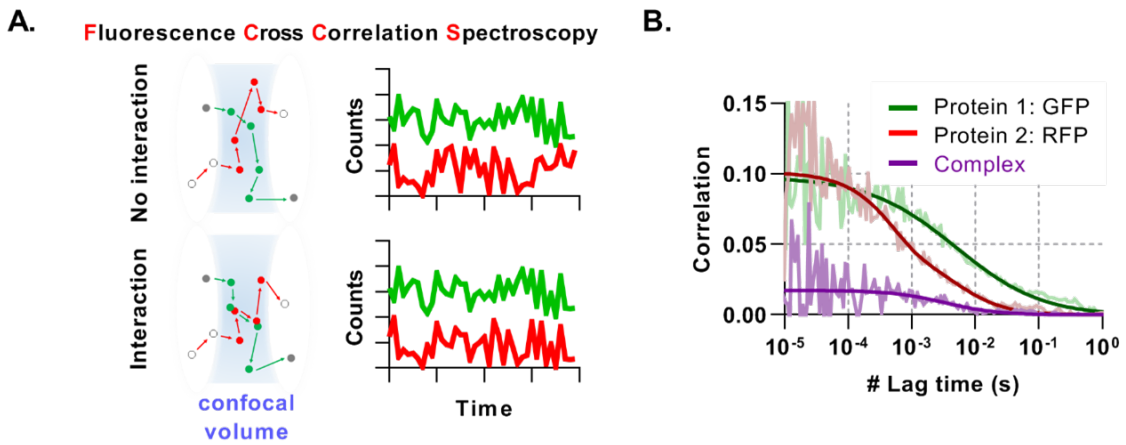

Figure 1.11: Fluorescence cross-correlation spectroscopy. (**A**) Schematic of green and red labelled fluorescent molecules diffusing in a small illuminated confocal volume (shaded) with the resulting count trace to the right. When particles interact the green and red signals are similar. (**B**) Correlation curves for an interacting pair of proteins. As the proteins interact more the cross-correlation between the signals (purple) increases and becomes a greater fraction of the auto-correlated curves (green and red).

FCS can be extended to measure molecular interactions by performing FCS simultaneously on two interacting molecules labelled with different coloured fluorophores. This dual colour method is called fluorescence cross-correlation spectroscopy (FCCS) (Bacia et al., 2006), whereby two spectrally distinct fluorophores, typically green and red, are measured simultaneously and auto-correlated independently (Figure 1.11A). These two sets of photon counts are then cross-correlated according to the equation

$$G_{\text{Cross}} = \frac{\langle I_G(t) I_R(t) \rangle}{\langle I_G(t) \rangle \langle I_R(t) \rangle}, \tag{1.14}$$

where $I_G(t)$ and $I_R(t)$ are the intensity of green and red photons respectively at time t. This cross-correlation follows a similar pattern to the auto-correlation, i.e. a diminishing correlation over progressively longer lag times, see Figure 1.11B. The cross-correlation is greater if the two signals are closely matched, which in this instance indicates that the two different sets of particles are moving together, from which it can be concluded that the proteins have complexed (Figure 1.11A). Analysing the fraction of free to bound components can then be used to determine the dissociation constant as,

$$K_D = \frac{[G_{\text{Free}}][R_{\text{Free}}]}{[GR]} \tag{1.15}$$

where $[GR] = \frac{G_{\text{cross}}}{G_G(0) G_R(0) N_A V_{\text{eff}}}$ is the concentration of the complex of red and green fluorescent molecules. The concentrations of the free green and red complexes are given by $[G_{\text{Free}}] = \frac{1}{G_G(0) N_A V_{\text{eff}}}$ and $[R_{\text{Free}}] = \frac{1}{G_R(0) N_A V_{\text{eff}}}$ respectively (Sadaie et al., 2014). Here, N_A represents the Avogadro constant and the effective confocal volume is given by $V_{\text{eff}} = \pi^{\frac{3}{2}} w_{xy}^2 w_z$. By plotting multiple repeat measurements of $[G_{\text{Free}}][R_{\text{Free}}]$ on the vertical axis against the $[GR]$ the horizontal axis, the K_D can be found from the slope of a line fitted to this data.

1.5 Aims and objectives

Daily rhythmic processes regulate many cellular, physiological, and behavioural processes in mammals, with their dysfunction leading to adverse health outcomes. Despite clear experimental determination of key molecular components of the mammalian circadian clock through genetic knock-out and mutation experiments, a comprehensive quantitative understanding of the relative contribution, role, and importance of these components is yet to be defined. Until recently the necessary tools to quantify the clock have been incomplete, namely knock-in fluorescent protein tags at the endogenous circadian gene loci via CRISPR/Cas9 and sensitive microscopy techniques necessary to quantify them. Live-cell imaging of proteins in time and space, alongside mathematical modelling, can aid our understanding by elucidating the critical rate-limiting constituents and mechanisms. Thus, predictive models informed by experimental data on top of a solid qualitative understanding of network architecture, can deepen understanding and inform future experiments through hypothesis generation. Furthermore, modelling may uncover experimentally inaccessible parameters and molecular species, such as the number of DNA bound molecules and protein-protein complexes. In this project I sought to improve and apply quantitative confocal imaging techniques to measuring key clock components to inform predictive mathematical models of the circadian clock. Specifically, my aims are to:

1. Quantify the abundance of, and interactions between, key circadian clock proteins comprising the core molecular circadian clock
2. Mathematically model the circadian clock to understand design principles behind the transcription-translation feedback loop driving oscillations, to generate new hypotheses and predictions of molecular behaviour
3. Develop and implement analytical tools for FCS and FCCS to improve techniques for quantifying fluorescently labelled proteins

1.6 Structure

This thesis is presented in the alternative (journal) format and thus three out of the four main results chapters (2-4) are presented as scientific articles that are either in-preparation or published following peer-review. The final chapter is presented as a traditional body of work that has not been written up in the format of a paper. Specifically, the results chapters are:

Chapter 2 **Regulation of CLOCK:BMAL1 DNA binding**

Contains the article published in eLife *Quantification of protein abundance and interaction defines a mechanism for operation of the circadian clock* (Koch et al., 2022).

In this paper, we present the use of quantitative microscopy methods and mathematical modelling to analyse rhythmic changes in CLOCK:BMAL1-DNA binding and its regulation by key repressor proteins, CRY1 and PER2. The quantification of CLOCK and BMAL1 proteins through FCS showed that the number of DNA targets was approximately 2-10 times more than CLOCK:BMAL1, indicating that this crucial transcription factor must rapidly move from site to site. Modelling demonstrated that PER2 plays a crucial role in enhancing the mobility of CLOCK:BMAL1 and facilitates its binding to new DNA target sites. Additionally, the results from modeling revealed that the oscillation of central circadian proteins, including CLOCK:BMAL1, CRY1, and PER2, occurs within an optimal range of concentrations and interaction affinities to modulate CLOCK:BMAL1-DNA binding. A significant aspect of this publication is my custom FCS analysis pipeline, built on traditional time-lagged correlation analysis techniques, which plays a crucial role in providing insights into these interactions.

Chapter 3 **Protein-protein interactions in the circadian clock**

Contains the in preparation article *Anatomy of circadian clock protein complexes in live cells.*

In this article in-preparation, we utilise Fluorescence Cross-Correlation Spectroscopy (FCCS) to assess the dynamic interactions between 15 key circadian clock proteins in live cells. Our findings reveal that the PER2-CRY1 and CRY1-BMAL1 protein pairs are the strongest interactors in the circadian network, with dynamic interaction strengths that are conserved across various cell types and protein expression levels. To determine the direct and facilitated interactions, we introduced new modelling frameworks and found that the majority of proteins in the circadian network directly bind to one another. In some cases, PER2 and BMAL1 were identified as central facilitators, forming scaffolds for multi-protein complexes in the repressive phase of the circadian cycle. Our results were validated by measuring three-way interactions through FCCS, which showed that strong PER2-CRY interactions can strongly stabilise CRY proteins and maintain post-translational rhythms of CRY2, independent of transcription. This study provides novel insights into the dynamic interactions of various circadian protein components within living cells and highlights the pivotal role of key interactions in the circadian clock network.

Chapter 4 **Improving analysis fluorescence correlation spectroscopy**

Contains the in preparation article *Accelerated analysis of fluorescence correlation spectroscopy.*

In this chapter, I introduce new analytical methods for FCS. The limitations of traditional curve-fitting approaches in FCS analysis are addressed, and alternative methods based on probabilistic models of molecular diffusion and photon emissions are discussed. This culminates in an article, in preparation, that develops an approximate likelihood function for FCS measurements, which can be used to estimate molecular concentration and diffusion rate. The estimators obtained from maximising this likelihood function provide improved inferences of these parameters by increasing the density of the data. Furthermore, these inferences are based on raw data from FCS experiments in the form of photon arrival times, which results in a higher number of data points, from approximately 100 to approximately 100,000.

Chapter 5 **Designing biological networks with ease: Network Designer**

In this concluding chapter, I introduce my new modelling tool, Network Designer. This application provides a user-friendly graphical interface for modeling and generating associated equations to facilitate rapid prototyping of various mathematical models. This significantly reduces the likelihood of committing errors commonly encountered when manually writing complex systems of coupled ordinary differential equations. Network Designer was utilised to generate models of crucial components of the circadian system, including the CLOCK:BMAL1 transcription factor binding to DNA in chapter 2 and protein-protein interactions leading to ternary complex formation in chapter 3. I elaborate on the design principles of Network Designer, provide illustrative examples of its application, and discuss how it can be adapted for future release.

Chapter 2

Regulation of CLOCK:BMAL1 DNA binding

2.1 Journal paper: Quantification of protein abundance and interaction defines a mechanism for operation of the circadian clock

Title Quantification of protein abundance and interaction defines a mechanism for operation of the circadian clock

Journal eLife

Year 2022

Authors **Alex A Koch***, James S Bagnall*, Nicola J Smyllie, Nicola Begley, Antony D Adamson, Jennifer L Fribourgh, David G Spiller, Qing-Jun Meng, Carrie L Partch, Korbinian Strimmer, Thomas A House, Michael H Hastings, Andrew SI Loudon

doi 10.7554/eLife.73976

* These authors contributed equally

2.1.1 My contributions

My contribution to this paper consisted of conceptualisation, data gathering, software, writing, reviewing, and editing. Alongside James Bagnall, I assisted with generation of fluorescent fusion cells and performed FCS experiments. I defined the mathematical models and I carried out all computational modelling work. Furthermore, I was responsible for programming the FCS analysis software with conceptual input from David Spiller, James Bagnall, Thomas House, and Korbinian Strimmer.

RESEARCH ARTICLE

Quantification of protein abundance and interaction defines a mechanism for operation of the circadian clock

Alex A Koch[1†], James S Bagnall[1†], Nicola J Smyllie[2], Nicola Begley[1], Antony D Adamson[1], Jennifer L Fribourgh[3], David G Spiller[1], Qing-Jun Meng[1], Carrie L Partch[3], Korbinian Strimmer[4], Thomas A House[4], Michael H Hastings[2], Andrew SI Loudon[1*‡]

[1]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom; [2]MRC Laboratory of Molecular Biology, Cambridge, United Kingdom; [3]Department of Chemistry and Biochemistry, University of California, Santa Cruz, Santa Cruz, United States; [4]Department of Mathematics, University of Manchester, Manchester, United Kingdom

Abstract The mammalian circadian clock exerts control of daily gene expression through cycles of DNA binding. Here, we develop a quantitative model of how a finite pool of BMAL1 protein can regulate thousands of target sites over daily time scales. We used quantitative imaging to track dynamic changes in endogenous labelled proteins across peripheral tissues and the SCN. We determine the contribution of multiple rhythmic processes coordinating BMAL1 DNA binding, including cycling molecular abundance, binding affinities, and repression. We find nuclear BMAL1 concentration determines corresponding CLOCK through heterodimerisation and define a DNA residence time of this complex. Repression of CLOCK:BMAL1 is achieved through rhythmic changes to BMAL1:CRY1 association and high-affinity interactions between PER2:CRY1 which mediates CLOCK:BMAL1 displacement from DNA. Finally, stochastic modelling reveals a dual role for PER:CRY complexes in which increasing concentrations of PER2:CRY1 promotes removal of BMAL1:-CLOCK from genes consequently enhancing ability to move to new target sites.

Editor's evaluation

The transcriptional negative feedback loop of the mammalian circadian clock is mainly regulated by interactions among BMAL1, CLOCK, PER1/2 and CRY1/2 in the nucleus. While the binding of CRY with BMAL1:CLOCK is known to block the transcriptional activity of BMAL1:CLOCK and the binding of PER:CRY dissociates BMAL1:CLOCK from DNA have been known, our understanding is limited in qualitative level. Koch et al., quantified the dynamic interactions among the core clock molecules such as their diffusion coefficients, binding affinity, and abundances in the nucleus. This greatly improves our understanding of the mammalian circadian clock. Importantly, this dynamic information is incorporated via a mathematical model to understand BMAL1-CLOCK binding to the target site (e.g., circadian proteins operate within an optimal range to modulate E-box binding), providing a coherent view on the mechanism driving the oscillation.

***For correspondence:** andrew.loudon@manchester.ac.uk

[†]These authors contributed equally to this work

Present address: [‡]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom

Competing interest: The authors declare that no competing interests exist.

Funding: See page 25

Preprinted: 28 August 2021
Received: 16 September 2021
Accepted: 11 March 2022
Published: 14 March 2022

Reviewing Editor: Naama Barkai, Weizmann Institute of Science, Israel

Introduction

The 24 hr light-dark cycles inherent to our planet have led to the evolution of molecular circuits capable of conveying time of day information, commonly known as circadian clocks. In mammals,

cell-autonomous circadian clocks operate in virtually all cells across tissues and enables coordination of numerous biological processes, including metabolism, immunity, and cell cycle progression (**Bhadra et al., 2017**; **Gibbs et al., 2014**). Autonomous cellular clocks are characterised by transcription-translation feedback loops (TTFLs), leading to cycles in protein and mRNA tuned to the 24 hr rhythms of the day-night cycle. Central to the mammalian circadian clock is the heterodimeric transcription factor comprised of CLOCK (circadian locomotor output cycles protein kaput) and BMAL1 (brain and muscle ARNT-like 1) that searches the genome to bind consensus sequence E-box sites (CANNTG), inducing expression of several hundred clock-controlled output genes every day. Targets include key circadian negative feedback regulators, Period (Per1, Per2, Per3), Cryptochrome (Cry1 and Cry2) and a secondary loop regulated by Nr1d1 and Nr1d2 (**Buhr and Takahashi, 2013**; **Gekakis et al., 1998**; **Huang et al., 2012**; **Liu et al., 2008**). These proteins act to repress the activity of CLOCK:BMAL1 to form a delayed negative feedback loop driving daily oscillations. In a current model, it is proposed that PER and CRY proteins dimerise to form a repressive complex with CK1 (casein kinase 1) to promote the removal of CLOCK:BMAL1 from DNA and thereby repress transactivation of target genes, while CRY1 independently binds the PAS domain core of CLOCK:BMAL1 and the BMAL1 transactivation domain leaving DNA binding intact whilst repressing recruitment of additional transcriptional coactivators (**Chiou et al., 2016**; **Xu et al., 2015**). An additional feedback loop is conferred by the protein REV-ERBα, which operates as a transcriptional repressor of Bmal1, resulting in a cycle of BMAL1 protein abundance (**Liu et al., 2008**).

Ultimately, a prerequisite for generation and output of cellular circadian rhythms is the ability of a finite pool of CLOCK:BMAL1 heterodimer protein to bind rhythmically to specific target sequences leading to the regulation of circadian gene expression in cells (**Koike et al., 2012**). Currently, we have very little insight into the quantitative biology of this process. Heterodimeric formation of transactivating and repressive complexes is a well-defined feature of the circadian molecular circuit, including the formation of CLOCK:BMAL1 and PER:CRY complexes (**Chiou et al., 2016**; **Huang et al., 2012**; **Xu et al., 2015**). Recently, PER:CRY proteins have been described as part of very large macromolecular complexes within the cell (**Aryal et al., 2017**). We have previously visualised several core circadian proteins, and from this measured the spatiotemporal profile and protein abundance for BMAL1 and PER2 (**Smyllie et al., 2016**; **Yang et al., 2020**). PER2 was found to cycle with a maximum amplitude of 12,000 copies per cell in fibroblasts and without circadian gating of nuclear localisation, contrary to observations in the *Drosophila* clock (**Shafer et al., 2002**; **Smyllie et al., 2016**). Only a relatively small amount of CRY1 is needed to localise PER2 to the nucleus, as shown in live SCN studies, with PER2 localisation remaining predominantly nuclear throughout the day (**Smyllie et al., 2022**). A recent study using a cancer cell line model has also shown that CRY1 protein remains nuclear at all circadian phases and at markedly higher abundance than its partner protein PER2 (**Gabriel et al., 2021**). In order to gain insight into the operation of core circadian clock proteins, we generated a genetically modified mouse in which CRY1 has been C-terminally fused with a fluorescent protein. We crossed this line to a previously described strain of mice expressing fluorescent-tagged BMAL1. We then used advanced imaging in both ectopically transformed cell lines and endogenously modified mice to characterise governing parameters in the regulation of CLOCK:BMAL1 DNA binding, including repression by PER2 and CRY1. We constructed mathematical models of the complex interactions and phased timings from multiple molecular species and experimentally inaccessible complexes, demonstrating how DNA binding in the peripheral circadian clock is regulated.

Using a combination of mathematical modelling and experimental validation, our data reveal that high-affinity interactions between circadian protein complexes serve to offset the low abundances of circadian proteins. In this way, the abundance of key components of the molecular clockwork is positioned optimally to regulate E-box binding. This is partly facilitated through PER2:CRY1 mediated displacement of CLOCK:BMAL1, such that PER2 protein serves a dual role, acting as both a component of the negative feedback arm but also to redistribute CLOCK:BMAL1 to new target sites. Thus, the stochiometric balance of PER:CRY with CLOCK:BMAL1 is critical for the elucidation of the full cellular circadian repertoire.

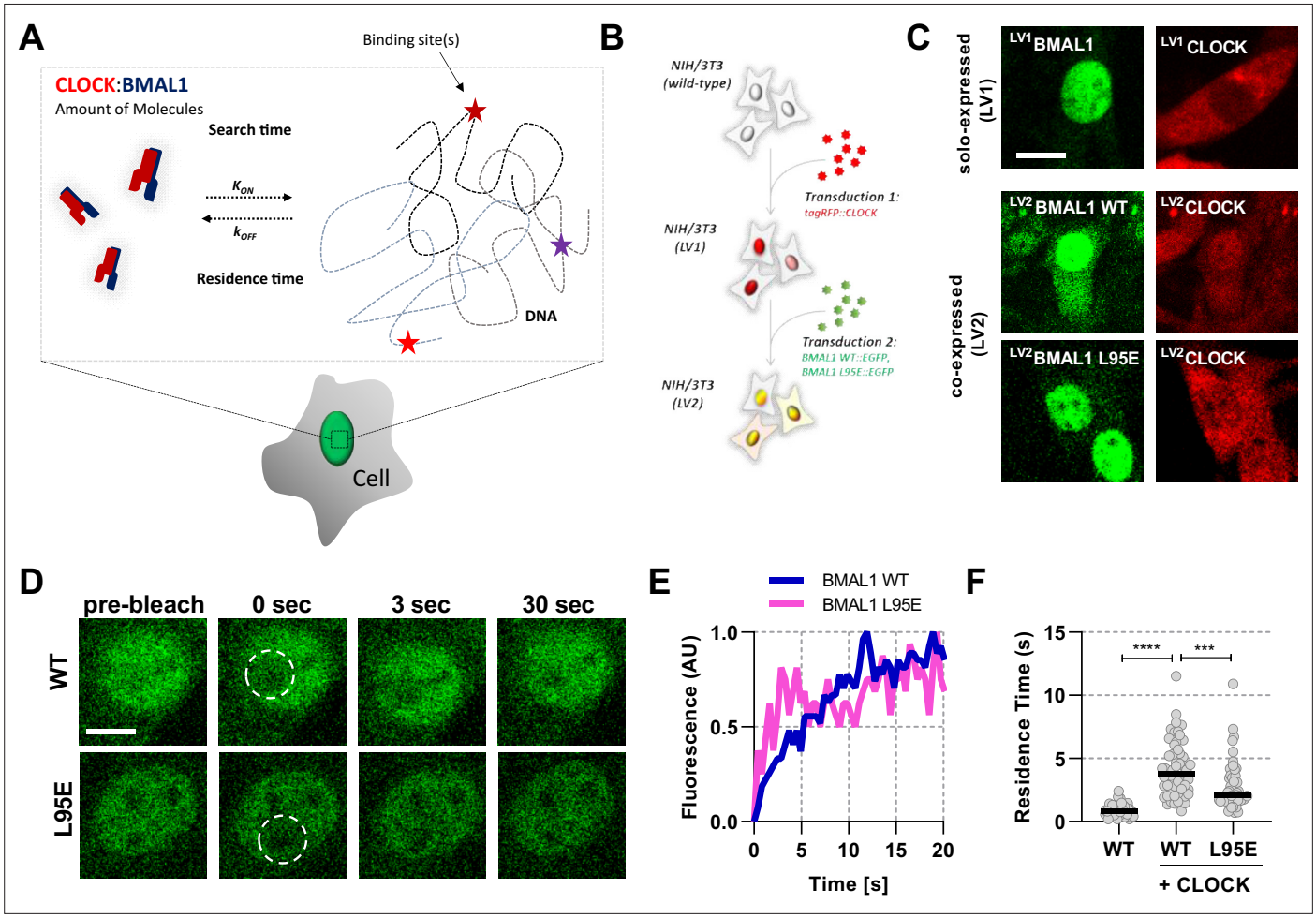

Figure 1. Short-lived DNA binding of BMAL1 and CLOCK. (**A**) Schematic representation of parameters regulating CLOCK:BMAL1 dimers binding to target DNA sites. (**B**) NIH/3T3 cells are either singularly or sequentially transduced to express fluorescent fusions with CLOCK or BMAL1 (wildtype and mutant variants).(**C**) Confocal microscopy images of cells solo-expressing (LV1) either tagRFP::CLOCK or BMAL1::EGFP or co-expressing (LV2) them together (including BMAL1 L95E DNA-binding mutant). (**D**) Confocal microscopy images for photobleaching of ^{LV2}BMAL1::EGFP-RFP::CLOCK labelled cells, either with wild-type or BMAL1 L95E DNA binding mutant. Images show nuclei and highlight region of bleaching. (**E**) Representative fluorescence recovery curves of bleach region for B. following normalisation. (**F**) Residence time calculated as the inverse of kOFF (s^{-1}), determined from fitting the recovery data with a single component binding model (n = 69, 58, and 51 cells). Bar represents median values. Source data for panel F available as ***Figure 1—source data 1***.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Source data 1. BMAL1 residence times.

Figure supplement 1. Ectopically expressed mRNA is the major form in a lentivirus transduced system.

Figure supplement 1—source data 1. Summary statistics.

Figure supplement 2. Binding plays a significant role in BMAL1 mobility.

Results

BMAL1 determines nuclear localisation and mobility of CLOCK

To quantify the properties of BMAL1 and CLOCK proteins (***Figure 1A***), we first used NIH/3T3 fibroblasts expressing fluorescent fusion proteins via a ubiquitin ligase C promoter, delivered by lentiviral transduction either singly (LV1) or as two sequential transductions (LV2) (***Figure 1B***; ***Bagnall et al., 2015***). Expression of the transgene was in >10 fold excess over the native unfused protein, as determined by single molecule Fluorescence In Situ Hybridisation (***Figure 1—figure supplement 1B***). Confocal microscopy of tagRFP::CLOCK or BMAL1::EGFP showed BMAL1 expression to be strongly

localised to the nucleus, whereas CLOCK was predominantly cytoplasmic when expressed alone (***Figure 1C***). Co-expression of both proteins in the same cells caused localisation of tagRFP::CLOCK to move to the nucleus, in agreement with earlier studies which showed cytoplasmic CLOCK localisation in BMAL1-deficient cells, and that circadian regulation of nuclear localisation of CLOCK correlated with BMAL1 availability (***Kondratov et al., 2003***; ***Stratmann et al., 2012***). We also transduced cells with a fluorescent fusion of a DNA-binding mutant of BMAL1, in which a leucine is substituted for glutamic acid in the basic helix-loop-helix region of the protein; referred to as L95E. The mutant BMAL1 also re-localised tagRFP::CLOCK protein to the nucleus from the cytoplasm in an manner equivalent to WT BMAL1 (***Figure 1C***; ***Huang et al., 2012***). We next performed Fluorescence Recovery After Photobleaching (FRAP) experiments to test the impact of CLOCK on the recovery dynamics of a bleached nuclear region of BMAL1::EGFP, by comparing responses with or without co-expressed tagRFP::CLOCK (***Figure 1D–F***). BMAL1 recovery half-life was found to be insensitive to the diameter of the bleached region, indicating that binding contributes to the recovery profile rather than this being a solely diffusion-led process (***Figure 1—figure supplement 2***; ***Sprague and McNally, 2005***). Reaction binding equations were fitted to determine the rate of dissociation, k_{OFF}, for BMAL1::EGFP, the reciprocal of which equates to an average characteristic duration of binding or residence time. Residence time of BMAL1::EGFP was increased in the presence of tagRFP::CLOCK ($p < 0.0001$), consistent with a requirement for CLOCK to bind DNA (***Figure 1F***). The mean residence time for the fluorescent CLOCK:BMAL1 complex was 4.13 s (95% CI, 0.57), a value consistent with DNA residence times for similar transcription factors (***Hettich and Gebhardt, 2018***; ***Stratmann et al., 2012***). Using the L95E DNA-binding mutant protein, we saw significantly reduced residence time of 2.83 s (95% CI, 0.54; $p = 0.0002$). Notably, the initial publication of the BMAL1 L95E mutant showed a twofold reduction in PER2::LUC expression and so suggests a strong relationship between DNA binding and transcriptional output (***Huang et al., 2012***).

To investigate this further we used Fluorescence (Cross) Correlation Spectroscopy (F(C)CS) (***Yu et al., 2021***), a technique used to determine live-cell concentration and diffusion properties of individually fluorescent-labeled BMAL1 and CLOCK proteins, as well as their interactions when co-expressed (***Figure 2A***). A normal diffusion model fitted the majority of data collected from cells expressing free EGFP or nuclear only NLS::EGFP proteins, as previously reported (***Dross et al., 2009***), whereas anomalous diffusion models – sub-diffusion caused by a range of factors such as DNA interactions and molecular crowding – accounted for a > 20% fraction, which in this instance may be explained by molecular crowding (***Tsekouras et al., 2015***). In comparison, for the fusion proteins, anomalous diffusion models accounted for >80% of all BMAL1 data sets (***Figure 2—figure supplement 1***). We used an anomalous diffusion model for all further analyses of circadian proteins to calculate diffusion coefficients and protein concentrations.

Singly expressed fluorescent BMAL1 and CLOCK were found to diffuse rapidly with a median coefficient of 9.2 $\mathrm{m^2\,s^{-1}}$ (SD, 3.3) and 12.6 $\mathrm{m^2\,s^{-1}}$ (SD, 6.1), respectively. In contrast, co-expression significantly reduced the rate of diffusion to 1.9 $\mathrm{m^2\,s^{-1}}$ (SD, 1.3; $p < 0.0001$) and 4.7 $\mathrm{m^2\,s^{-1}}$ (SD, 3.2; $p < 0.0001$) for BMAL1 and CLOCK ,respectively (***Figure 2B***). The L95E mutant diffused more rapidly than WT BMAL1, consistent with fewer interactions with DNA in the nucleus (***Figure 2C***). When co-expressed, WT BMAL1 and CLOCK exhibited a 2:1 concentration ratio in the nucleus (***Figure 2D***, ***Figure 2—figure supplement 2E***), presumably arising from a combination of differences in protein turnover, shuttling and direct interaction. F(C)CS was then used to observe this interaction and determine a live-cell dissociation constant (K_D; reciprocal measure of affinity) (***Krieger et al., 2015***). A positive cross correlation curve was observed between BMAL1::EGFP and tagRFP::CLOCK that was not apparent in cells expressing NLS::EGFP with tagRFP::CLOCK (***Figure 2E***). To calculate K_D, we fitted a one-site saturating binding curve to the relationship between heterodimer and monomer which yielded a value of 148 nm (SD, 9.8) for WT BMAL1::EGFP and tagRFP::CLOCK (***Figure 2F***). The K_D was measured for cells with the reverse fluorescent protein labelling, namely EGFP::CLOCK and BMAL1::-tagRFP, finding similar a value of 145 nm (SD, 4.8), although a stronger interaction was found in vitro by surface plasmon resonance (***Figure 2—figure supplement 2F***). Previous work found that the V435R mutation of BMAL1 in the PAS-B domain, leads to reduced dimerisation with CLOCK (***Huang et al., 2012***). We used the V435R mutation to confirm our F(C)CS measurements by co-expressing V435R-BMAL1 and WT-CLOCK. This elicited a ≈1.5-fold reduction in interaction affinity, resulting in a K_D of 201 nm (SD, 14) (***Figure 2G***) and a reduction from 2:1 to a 4:1 ratio of BMAL1 and CLOCK in

Figure 2. Live-cell interaction measurements demonstrate BMAL1 and CLOCK mobility is regulated by dimerisation and DNA binding. (**A**) Schematic of confocal volume used in FCCS with corresponding photon count traces. Interaction may be seen by correlation between both channels. Representative auto- and cross- correlation data showing raw data and fit lines for monomeric and complexed fluorescent proteins. (**B**) FCS data showing diffusion for BMAL1 and CLOCK in solo- and co-expressed conditions (n = 173, 152, 198, and 185 cells). (**C**) FCS results for BMAL1::EGFP diffusion for NIH/3T3 cells that co-express tagRFP::CLOCK. Data shown is for comparison of BMAL1 as either wild-type of L95E DNA-binding mutant. Bars show median and interquartile range. (**D**) Correlation of nuclear protein quantification showing relationship of BMAL1::EGFP with tagRFP::CLOCK for both wildtype and DNA binding mutant (n = 221 cells from three biological replicates). (**E**) Average cross-correlation curves for BMAL1::EGFP (WT) with tagRFP::CLOCK (n = 140) compared to a non-interacting control of NLS::EGFP co-expressed with tagRFP::CLOCK (n = 408). (**F**) Dissociation plot from FCCS data for BMAL1::WT and tagRFP::CLOCK. (**G**) Summary of calculated dissociation constants across all conditions, including BMAL1 dimerisation mutant, V435R

Figure 2 continued on next page

Figure 2 continued

(n = 156, 274, and 244). Mann-Whitney non-parametric test to determine significance (values are denoted as p > 0.05 ns, p < 0.05 *, p < 0.01 **, p < 0.001 *** and p < 0.0001 ****). Source data for panels B,C available as ***Figure 2—source data 1*** and panel E as ***Figure 2—source data 2***.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. BMAL1 and CLOCK FCS diffusion rates.

Source data 2. BMAL1 and CLOCK paired FCS concentrations.

Figure supplement 1. Anomalous diffusion best fits protein movement.

Figure supplement 2. Fluorescent BMAL1 and CLOCK proteins behave similarly when colours are swapped.

the nucleus (***Figure 2—figure supplement 2E***). In contrast, the BMAL1 L95E DNA binding mutant showed no difference in interaction affinity compared to WT BMAL1 protein. These data demonstrate that BMAL1 is a critical determinant of the localisation, mobility and concentration of CLOCK in the nucleus.

From this, we can infer the abundance of nuclear CLOCK from measurements of BMAL1, and make use of available endogenously labelled Venus::BMAL1 mice to measure remaining DNA binding parameters. First, we confirmed our cell line measurements for binding rates and diffusion using the Venus::BMAL1 mice (***Yang et al., 2020***), finding that they remain within a similar range across a number of primary cell types, including macrophages and pulmonary fibroblasts (***Figure 3—figure supplement 1***). We also measured protein number of the endogenous BMAL1, observing that total copies per nucleus vary from 1000 to 10,000 between individual cells, likely due to desynchrony and differing nuclear volumes. Moreover, a large overlap in nuclear copy numbers was observed across all cell types despite substantial changes in the mean. These data are critical in our understanding of the ratio of BMAL1 to target sites, effectively determining the capacity to regulate the full repertoire of target genes within a specific cell type.

Quantification of strong rhythmic interaction of BMAL1 with CRY1

The ability to measure BMAL1 amounts to infer copy number of CLOCK, allows us to measure other critical pairings with BMAL1. This includes the repressive action of CRY1 binding to CLOCK:BMAL1, resulting in reduced transactivation. To explore the interaction between BMAL1 and CRY1, we generated a genetically modified mouse in which CRY1 has been C-terminally fused with the red fluorescent protein mRuby3 using CRISPR-mediated genomic editing to insert the coding sequence, replacing the endogenous CRY1 stop codon (***Figure 3—figure supplement 3A***; ***Bajar et al., 2016***; ***Bennett et al., 2021***). First, to test any potential impact on circadian pace-making, we measured CRY1::mRuby3 fluorescence in whole-field organotypic SCN slices (***Figure 3—figure supplement 3B***) which exhibited regular cycles in red fluorescence with a period of 23.9 hr (SD, 0.6) (***Figure 3—figure supplement 3C***; ***Smyllie et al., 2016***). Additionally, wheel running measurements of these mice confirmed normal behavioural rhythmicity (***Figure 3—figure supplement 2***). We next crossed these mice to the Venus::BMAL1 mouse line (***Yang et al., 2020***), previously inter-crossed with a PER2::LUCIFERASE background (***Yoo et al., 2004***) to provide an independent circadian phase-reference marker (referred to as BMAL1xCRY1 labelled mouse). Using isolated lung fibroblasts from BMAL1xCRY1 mice we assessed bioluminescence in response to dexamethasone (DEX) synchronisation, and observed 23.3 hr cycles (SD, 0.6) which were sustained for >4 days (***Figure 3—figure supplement Figure 3—figure supplement 3D, E***). From this, we are confident that the fluorescent fusion proteins do not disrupt the normal operation of the circadian pacemaker.

Using the same synchronisation approach, we then measured BMAL1xCRY1 fluorescence in single cells every 4 hr from 24 to 48 hr post-DEX synchronisation, using F(C)CS (***Figure 3A-B***). Both fluorescent signals were localised to the nucleus. Venus::BMAL1 showed a consistent diffusion pattern over a circadian cycle, with a mean diffusion coefficient of 0.58 $m^2 s^{-1}$ (SD, 0.03), whereas CRY1 mobility exhibited circadian variance, with slow diffusion 28 hr post-DEX and elevated diffusion rates 12 hr later (***Figure 3C***). Interestingly, this change in mobility is consistent with a binding to a mass equivalent to the molecular weight of PERIOD2. Peak protein concentrations of BMAL1 and CRY1 had an approximate and appropriate phase-separation of 8 hr (***Fustin et al., 2009***). Auto-correlation analyses

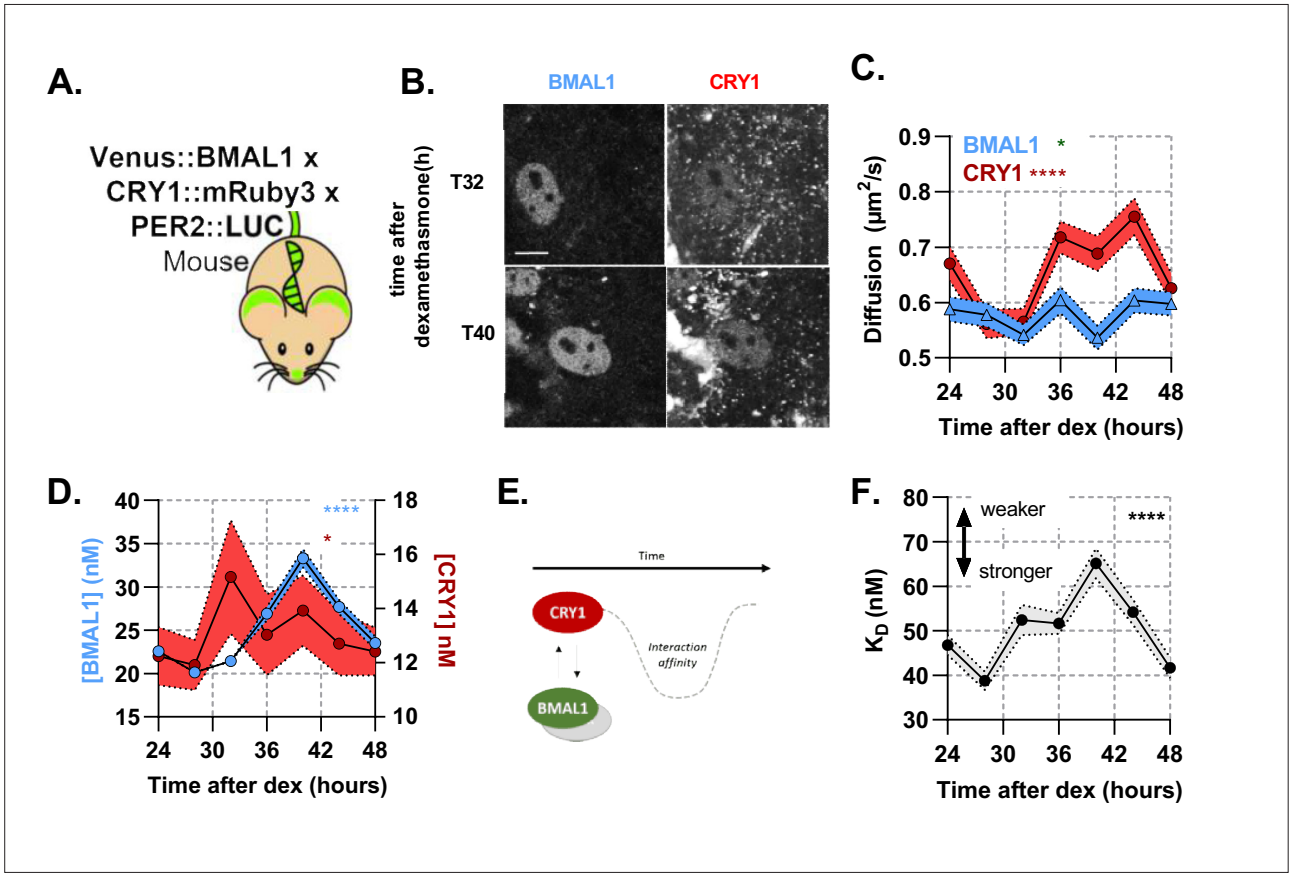

Figure 3. A rhythmic and strong interaction observed between slow-diffusing BMAL1 and CRY1 facilitates repression. (**A**) Schematic of triple-labelled mice from which primary lung fibroblasts were isolated (**B**) Confocal images of two cells shown for Venus::BMAL1 and CRY1::mRuby3 over time. FCS determined measurement for diffusion coefficient (**C**) and protein concentration (**D**) of Venus::BMAL1 and CRY1::mRuby3 (n = 136, 143, 173, 131, 158, 121, and 132; line shows the mean and error envelopes show the SEM). (**E–F**) Interaction strength between BMAL1 and CRY1 was also measured over time as illustrated by the schematic of affinity as well as plotted values of dissociation constant (error envelope shows the standard deviation). Kruskal-Wallis test used to determine significance (values are denoted as $p > 0.05$ ns, $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ *** and $p < 0.0001$ ****). Source data for panels B,C available as ***Figure 3—source data 1***, ***Figure 3—source data 2***, ***Figure 3—source data 3***, ***Figure 3—source data 4***.

The online version of this article includes the following source data and figure supplement(s) for figure 3:

Source data 1. CRY1 FCS diffusion rates.

Source data 2. BMAL1 FCS diffusion rates.

Source data 3. BMAL1 FCS concentration.

Source data 4. CRY1 FCS concentration.

Figure supplement 1. BMAL1 concentration and DNA binding parameters minimally vary across cell types.

Figure supplement 1—source data 1. CRY1::mRuby3 mouse genotyping.

Figure supplement 2. Generation of CRY1::mRuby3 mouse line.

Figure supplement 3. Triple endogenous labelled mice used to assay rhythms in SCN and peripheral lung fibroblasts.

revealed the concentration of BMAL1 is on average 1.92 fold (SD, 0.32) higher than CRY1, with a mean concentration of 29.3 nm and 13.4 nm respectively (equating to approximately 16,000 and 7000 molecules per nucleus), consistent with the range we reported earlier for PER2 (***Smyllie et al., 2016***). The amplitude of CRY1 was found to be shallow, cycling from 11.9 nm (SD, 5.7) at T28 to 15.2 nm (SD, 14.0) at T32, comparable to the approx. 25% amplitude observed for CRY1 in the SCN (***Figure 3—figure supplement 3C***). BMAL1 demonstrated a larger amplitude, cycling from 20.1 nm (SD, 7.1) at T28 to a peak of 33.3 nm (SD, 13.6) 40 hr after DEX (***Figure 3D***).

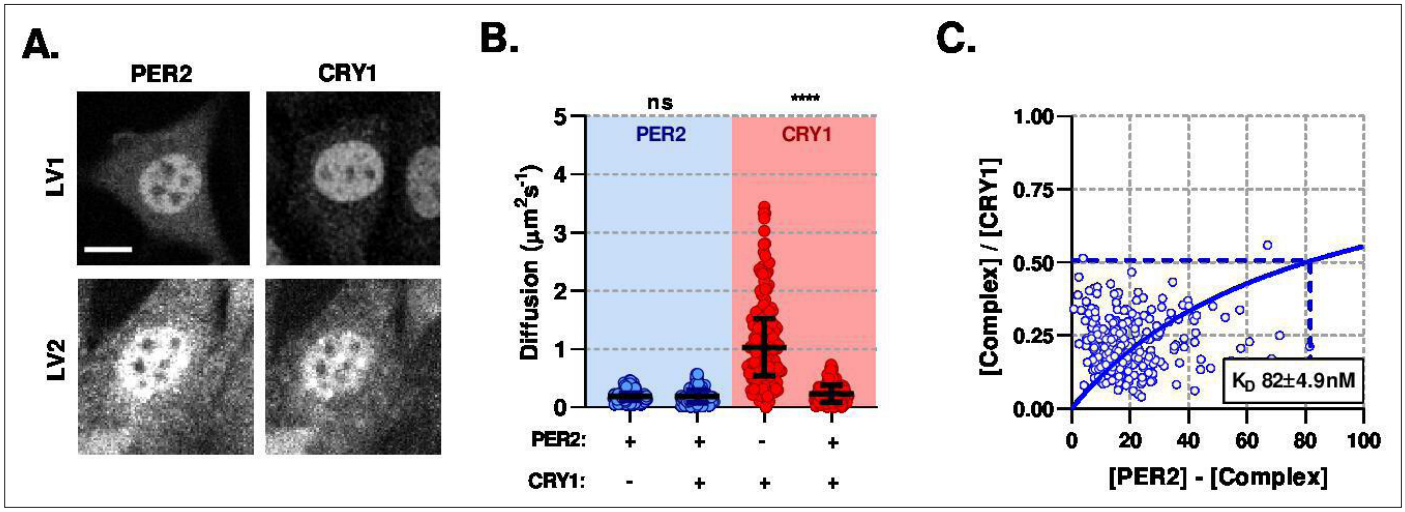

Figure 4. PER2 modulates CRY1 mobility via a high-affinity association. (**A**) Confocal images of transduced NIH/3T3 cells that either solo- or co- express PER2 and CRY1. (**B**) FCS data showing diffusion for PER2 and CRY1 in solo- and co-expressed conditions (n = 165, 174, 274, and 274 cells; diffusion rate means of 0.2, 0.2, 1.1, 0.2). (**C**) Dissociation plot from nuclear FCS measurements for EGFP::PER2 and CRY1::tagRFP (n = 274). Significance determined by Mann-Whitney test (values are denoted as p > 0.05 ns, p < 0.05 *, p < 0.01 **, p < 0.001 *** and p < 0.0001 ****). Source data for panels B available as ***Figure 4—source data 1***.

The online version of this article includes the following source data and figure supplement(s) for figure 4:

Source data 1. PER2 and CRY1 FCS diffusion rates.

Figure supplement 1. CRY1 mobility is affected by co-expression with PER2.

We then analysed the interaction affinity between BMAL1 and CRY1 over time (***Figure 3F***). This interaction exhibited significant changes over a 24 hr cycle (p = < 0.0001), with the strongest interaction at T28, K_D = 38.8 nm (SD, 2.1), and weakest at T40, K_D = 65.1 nm (SD, 3.4) (***Figure 3F***). These profiles were found to correlate with the diffusion profile of CRY1 (***Figure 3C***). Intriguingly, the mean interaction strength between BMAL1 and CRY1 is >2 fold stronger than that between BMAL1 and CLOCK (***Figure 2G***). A similar relationship was found in vitro when measuring interactions using biolayer interferometry (***Fribourgh et al., 2020***). Although, these interaction measurements do not distinguish between whether either proteins are complexed with other partners, diffusion data is consistent with BMAL1 being bound to CLOCK, and are compatible with a model in which the low abundance of the CRY1 repressor is offset by a high affinity for the CLOCK:BMAL1 heterodimer.

The changes in the diffusion profile of CRY1 are consistent with its association with an additional binding partner, such as PER2, thereby altering the affinity of CRY1 for CLOCK:BMAL1 (***Fribourgh et al., 2020***; ***Ye et al., 2014***). To measure the interaction between CRY1 and PER2 directly, we transduced NIH/3T3 cells with lentivirus so that cells constitutively express EGFP::PER2 or CRY1::tagRFP fusion proteins. In both cases, the expressed protein was found to localise predominately to the nucleus, although some cytoplasmic fluorescence was observed. When co-expressed, subcellular localisation was unchanged, although large punctate aggregates of signal were observed (***Figure 4A***). PER2 was found to have the slowest diffusion coefficient recorded within all our F(C)CS measurements, when in the non-aggregate space. PER2 mobility was not altered following co-expression with CRY1, whereas CRY1 exhibited reduced mobility in the presence of ectopic EGFP::PER2 (***Figure 4B***). The diffusion coefficient for CRY1 co-expressed with PER2 was similar to measurements of the endogenous protein (***Figure 3C***), suggesting PER2 and CRY1 exhibit similar stoichiometry within these cells. The anomalous diffusion model fit the majority of data sets, including ^{LV1}CRY1, ^{LV1}PER2, and ^{LV2}PER2. However, normal diffusion models accounted for >50% of ^{LV2}CRY1 correlation analyses suggesting a distinct change in CRY1 following interaction with PER2 (P < 0.0001), potentially from a loss of significant DNA binding of the CLOCK:BMAL1 complex (***Figure 4—figure supplement 1***). Best fit models for each data set demonstrated a strong affinity between PER2 and CRY1 with a K_D of 81.8 nm (SD, 4.9) (***Figure 4C***) and consistent with previous in vitro measurements (***Schmalen et al., 2014***).

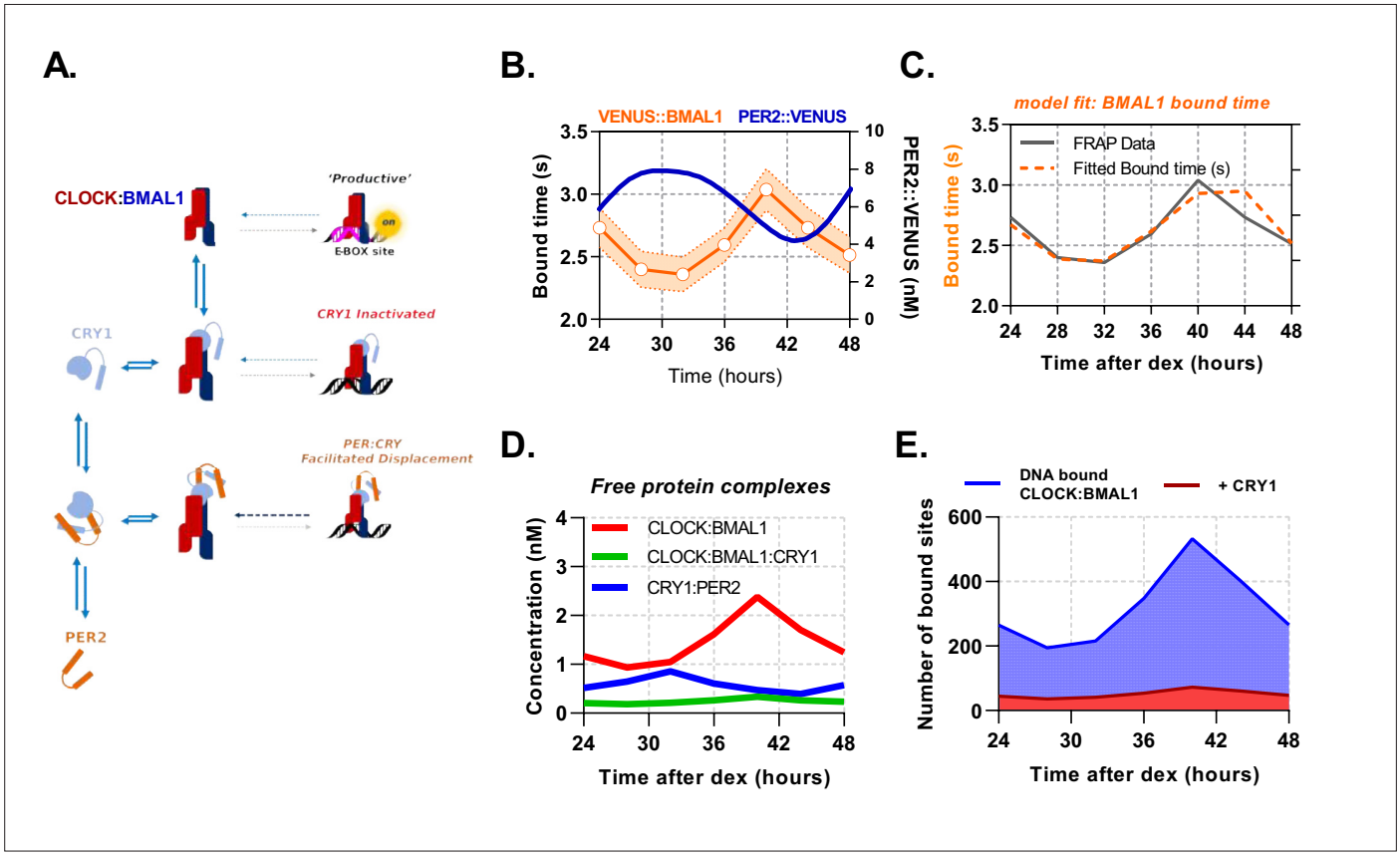

Figure 5. PER2 acts via CRY1 to mediate rhythmic displacement of CLOCK:BMAL1 from DNA. (**A**) Schematic representation of model topology used for the deterministic model of CLOCK:BMAL1 DNA binding. (**B**) Primary lung fibroblasts from BMAL1 x CRY1 x PER2 mice were synchronised with dexamethasone. Plot shows PER2 concentration as measured via FCS by ***Smyllie et al., 2016*** as well as mean BMAL1 binding time (showing SEM error envelope). Binding time was measured by confocal FRAP measurements performed on the Venus::BMAL1 fluorescence. Orange line shows the inverse of kOFF (s^{-1}), determined from fitting the recovery data with a single component model (n = 48, 70, 82, 63, 82, 64, and 65 cells). (**C**) ODE model was fit to FRAP binding data from E. and using a measured input for PER2 nuclear concentration previously determined in ***Smyllie et al., 2016***. Model output showing (**D**) inferred nuclear concentrations for molecular complexes (**E**) and CLOCK:BMAL1 without and with CRY1 bound to target sites (see supplementary materials for parameters). Panel B has been adapted from Figure 3C from ***Smyllie et al., 2016***.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. ODE model of CLOCK:BMAL1 DNA binding using measured inputs and modelled perturbations.

Quantitative data are an enabling and often essential component of stringent mathematical modelling of cell signalling (***Bagnall et al., 2018***). Having quantified the necessary parameters, we then sought to use them in developing a mathematical model of CLOCK:BMAL1 DNA binding, with the aim of understanding how the multiple regulatory motifs of changing molecular concentrations, interactions and binding kinetics coalesce to regulate DNA binding and transcriptional activation of BMAL1. We explored multiple topologies that were able to fit BMAL1 binding rhythms, carrying the simplest model forward that incorporated our measured data. We modelled the system using a set of ordinary differential equations (ODEs) to depict a current understanding of the system; BMAL1 dimerises with CLOCK, which may subsequently bind and unbind to DNA target sites. To model repression, CRY1 may either inactivate CLOCK:BMAL1 via direct binding or, via dimerisation to PER2, form PER2:CRY1 (mimicking complexes with CK1) to displace CLOCK:BMAL1 from DNA (***Figure 5A***; ***Chiou et al., 2016***; ***Huang et al., 2012***; ***Koike et al., 2012***; ***Xu et al., 2015***). The latter would presumably lead to rhythmic changes in the residence time of BMAL1 and provide a sensible option to fit and complete the model.

To assess dynamic changes in inferred DNA binding rates of BMAL1, we isolated lung fibroblasts from BMAL1xCRY1 mice and determined the k_{OFF} values by FRAP following DEX synchronisation.

Table 1. Summary of ordinary differential equation model parameters.
Model fit $\chi^2 = 7.46$.

Input parameters

Parameter	Unit	Description	Value±SD
K_D(C:B - E-Box)	nM	CLOCK:BMAL1 - E-Box dissociation constant	10 (***Huang et al., 2012***)
K_D(C-B)	nM	CLOCK - BMAL1 dissociation constant	147.6 ± 9.8
K_D(B-C1)	nM	BMAL1 - CRY1 dissociation constant	Time-point dependent, ***Figure 3F***
K_D(C1-P2)	nM	CRY1 - PER2 dissociation constant	81.8 ± 4.9

Fitted parameters

Parameter	Unit	Description	Value±SD (Inverse Hessian eigenvalue of fit)
k_{ON}	$\mathrm{nm}^{-1}\,\mathrm{s}^{-1}$	CLOCK:BMAL1 - DNA binding on rate	$0.027 \pm 1.034\ (1.96)$
d_{ON}	$\mathrm{nm}^{-1}\,\mathrm{s}^{-1}$	BMAL1 - CRY1 binding rate	$0.237 \pm 1.003\ (3.42 \times 10^{-2})$
a_{ON}	$\mathrm{nm}^{-1}\,\mathrm{s}^{-1}$	PER2 - CRY1 binding rate	$6.34 \pm 1.00\ (6.79 \times 10^{-8})$
R_{OFF}	s^{-1}	CLOCK:BMAL1:CRY1:PER2 - DNA unbinding rate	$(1.23 \pm 0.33) \times 10^{1}\ (1.00)$
b_{ON}	$\mathrm{nm}^{-1}\,\mathrm{s}^{-1}$	CLOCK - BMAL1 binding rate	$9.17 \pm 1.29\ (1.00)$

Derived parameters

Parameter	Unit	Description	Value±SD
k_{OFF}	s^{-1}	CLOCK:BMAL1 - DNA binding unbinding rate	$k_{\mathrm{ON}} \times K_D(\text{C:B - E-Box}) = 0.27 \pm 10.34$
b_{OFF}	s^{-1}	CLOCK - BMAL1 unbinding rate	$b_{\mathrm{ON}} \times K_D(\text{C-B}) = (1.35 \pm 0.21) \times 10^{3}$
d_{OFF}	$\mathrm{nm}^{-1}\,\mathrm{s}^{-1}$	BMAL1 - CRY1 unbinding rate	Time-point dependent, $d_{\mathrm{ON}} \times K_D(\text{B:C1})$
a_{OFF}	s^{-1}	PER2 - CRY1 unbinding rate	$a_{\mathrm{ON}} \times K_D(\text{C1-P2}) = (5.19 \pm 0.88) \times 10^{2}$

Measurements of BMAL1 protein recovery were taken every 4 hr from 24 hr to 48 hr post-DEX, showing k_{OFF} to be rhythmically regulated (***Figure 5B***). The BMAL1 k_{OFF} profile was in antiphase to recordings of nuclear PER2 concentrations from ***Smyllie et al., 2016*** (***Figure 5B***). To fit all parameters to the model (***Table 1***), the measured concentrations of PER2 (***Smyllie et al., 2016***), CRY1 and BMAL1 were used as inputs, using data described in (Figs. ***Figures 3D and 5C***). On/Off rates were constrained to measured dissociation constants from F(C)CS (***Table 1***), with the K_D value between BMAL1 and E-box sites set at 10 nm, as measured by ***Huang et al., 2012***. Using a mean value from multiple published ChIP-Seq data, the potential number of DNA target sites was set as 3,436 (***Table 2***).

The ODE model was then fitted by simulating FRAP so that a model-derived k_{OFF} could be used against our experimental data (***Figure 5B***) via Chi^2 minimisation (Chi^2, 7.46) to mean and standard error on the mean (***Figure 5C***, ***Table 1***). In order to confirm identifiability of the unknown parameters, we calculated the eigenvalues of the Hessian matrix of the fit, finding that it is non-singular and reasonably well conditioned (***Table 1***). Next, we used this model to infer

Table 2. BMAL1 ChIP reports.

No.	Tissue	BMAL1 peaks	Reference
1	Liver	2049	***Rey et al., 2011***
2	Liver	5952	***Koike et al., 2012***
3	U2OS	2001	***Wu et al., 2017***
4	PECS	2026	***Oishi et al., 2017***
5	Liver	4813	***Beytebiere et al., 2019***
6	Kidney	4034	***Beytebiere et al., 2019***
7	Heart	2520	***Beytebiere et al., 2019***
8	NIH3T3	4740	***Chiou et al., 2016***
9	Skeletal muscle	2787	***Dyar et al., 2018***
	Mean average	3436	

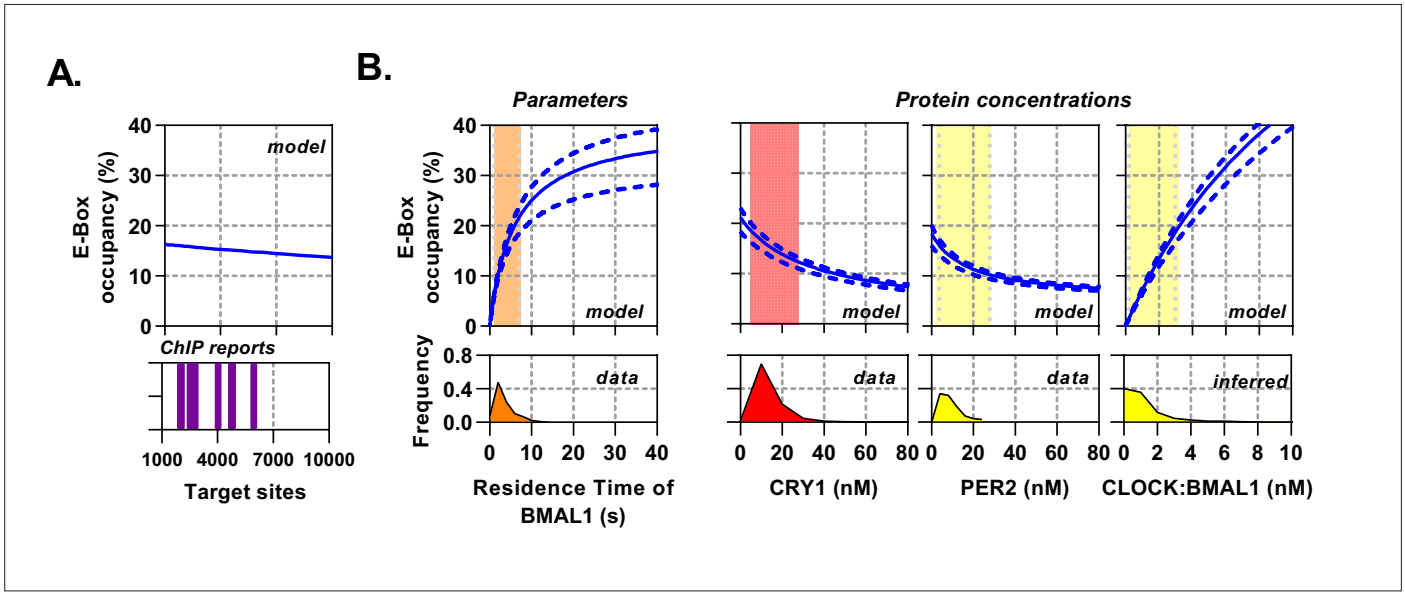

Figure 6. Circadian proteins operate within an optimal range to modulate E-Box binding. Sensitivity analysis of the deterministic binding model showing relationship of measured parameters (bottom) against model for occupancy of active BMAL1:CLOCK on target sites (top). (**A**) Changing number of target sites with data matched to BMAL1 ChIP data sets. (**B**) From left to right, the effect of changing residence time of CLOCK:BMAL1, or protein concentrations. Histograms show measured concentrations for corresponding proteins across all conditions/cells. The 10th to 90th percentile is highlighted. Source data available as ***Figure 6—source data 1***.

The online version of this article includes the following source data for figure 6:

Source data 1. Model OAT outputs.

experimentally inaccessible complexes, specifically PER2:CRY1, CLOCK:BMAL1, and CLOCK:BMAL1:CRY1 (***Figure 5D***). We find that free CLOCK:BMAL1 (unbound to DNA) cycles in remarkably low abundance from 0.9 nm to 2.4 nm, which equates to a change of ca. 809 molecules, similar to that of the PER2:CRY1 complex. Furthermore, predicted DNA binding of CLOCK:BMAL1 has an average baseline of 194 sites bound at any one time, rising to 526 sites at the peak, in agreement with the expected ≈2-fold peak enrichment from ChIP reports (***Beytebiere et al., 2019***; ***Koike et al., 2012***; ***Figure 5E***). The model also suggests ≈50% of the available transcription factor complex is engaged with site-specific interactions with availability predominantly limited by the K_D with CLOCK. Finally, DNA-bound CLOCK:BMAL1:CRY1 complex was persistent, with low abundance cycling from 32 to 69 target sites (accounting for 11% of total BMAL1 bound sites).

Circadian proteins are within an optimal expression range to modulate E-Box binding

The topology of the circadian molecular circuit is preserved across all cell types, yet it is also known that different cell types have widely differing repertoires of target genes and accessible genomic target sites for CLOCK:BMAL1 to bind (***Beytebiere et al., 2019***). We therefore pursued the extent to which varying the number of target sites may have an impact on the available pool of CLOCK:BMAL1 to bind target sequences, as calculated by site occupancy (the % sites occupied at any given moment). We simulated the model over a biological range of binding sites (1000 – 10,000), informed by multiple BMAL1 and CLOCK ChIP data sets (***Figure 6A***; ***Beytebiere et al., 2019***; ***Koike et al., 2012***). We found target site occupancy decreased marginally from 16.2% to 13.6%, showing that any variance between different numbers of available target sites has minimal impact. We then explored how varying binding parameters affected site occupancy, relating them to the variability observed in our data sets but considering values beyond these limits. The CLOCK:BMAL1 unbinding rate accounted for a 21.5% change when residence times across the observed physiological range are considered (***Figure 6B***); outside of this range occupancy begins to saturate so that a further 30 s increase in residence time only accounts for an additional 12.5% binding. Therefore, the unbinding rate, as measured

experimentally, is optimally positioned to regulate target site occupancy in a manner consistent with the displacement mechanism governed by PER:CRY.

Additionally, protein concentrations vary across circadian time as well as individual cells and cell types (***Figure 3—figure supplement 1***). We therefore simulated target-site occupancy across varied biologically plausible concentrations for CRY1, PER2, and CLOCK:BMAL1 and calculated the fraction of occupied sites (***Figure 6B***). Increasing CRY1 and PER2 led to a reduction in target-site occupancy, whereas a rise in CLOCK:BMAL1 led to a substantial increase and in both cases. Moreover, the biologically observed range occupied the most sensitive part of the curve, such that oscillations in protein copy number can evoke significant changes of occupancy. Hence, the system is positioned to make efficient use of the biological concentrations of the constituent proteins.

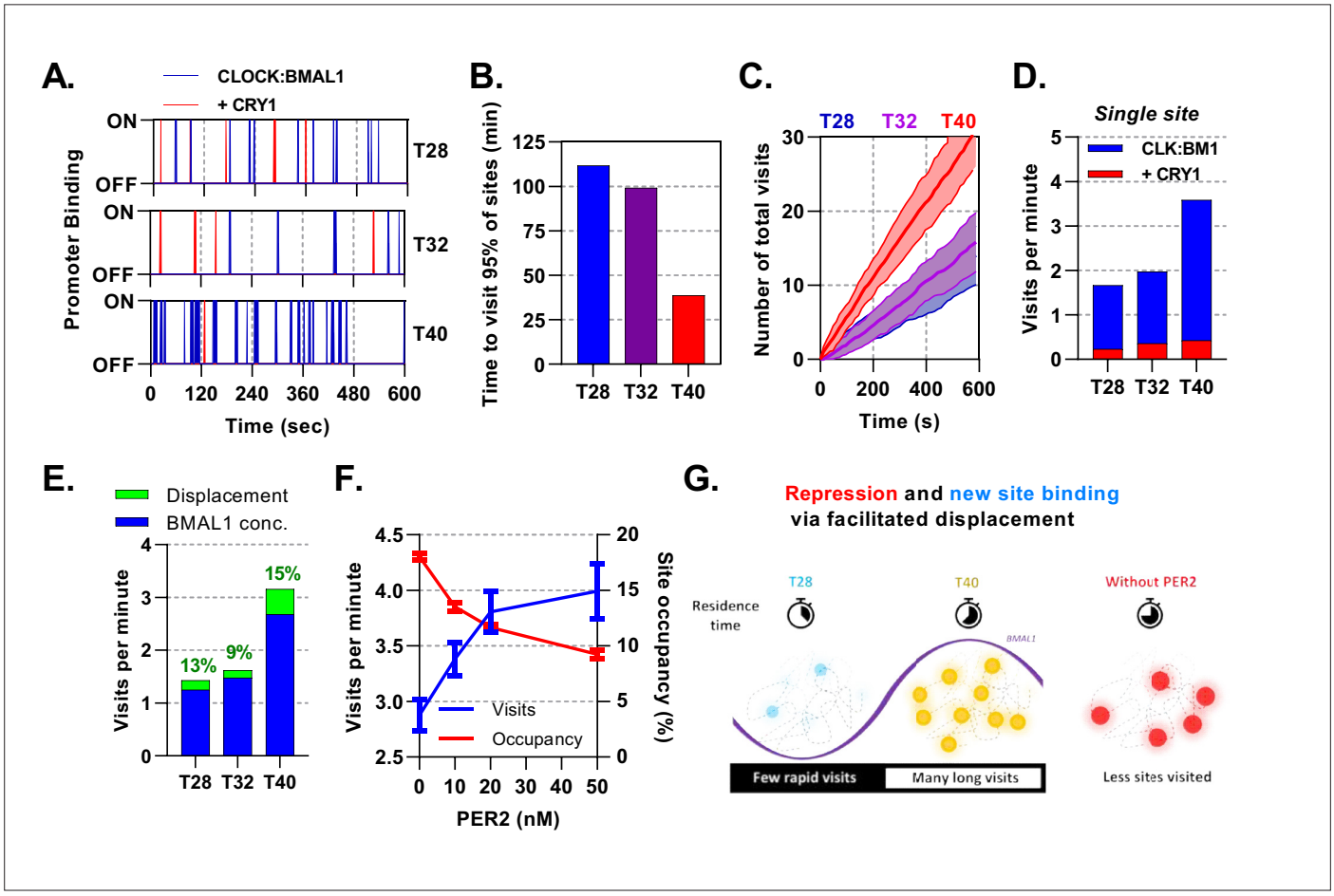

Figure 7. Mathematical modelling demonstrates dual function of PER:CRY mediated repression. Stochastic binding model outputs using parameters corresponding to T28, T32 or T40 post dexamethasone BMAL1 x CRY1 data sets. (**A**) Shows a promoter corresponding to the average binding rate of CLOCK:BMAL1, (**B**) the time to visit 95% of target sites once and (**C**) number of visits to a single promoter over time. Shaded error envelope shows standard deviation. (**D**) Average number of visits per minute to a target site showing active and CRY1 repressed CLOCK:BMAL1 visits. (**E**). Comparison of the contribution of BMAL1 concentration (blue) and PER2 facilitated displacement (green) on the visits per minute to a target site. Percentage contribution indicated. (**F**) Relationship of PER2 protein concentration to site visitations per minute and occupancy by CLOCK:BMAL1 using parameters for T40 explored over different concentrations of PER2. Error bars represent standard deviation. (**G**) The action of CRY:PER leads to short-lived transient binding of CLOCK:BMAL1 to DNA, working as both a repressive action whilst also facilitating binding to new target sites.

The online version of this article includes the following figure supplement(s) for figure 7:

Figure supplement 1. Stochastic binding model using experimentally measured parameters (**A**) Stochastic model showing the average binding (with SD) of CLOCK:BMAL1 bound target sites using input measurements from all time points for both WT and without PER2 simulations.

Mathematical modelling demonstrates dual function of PER2:CRY1 mediated repression

Site occupancy is a function of the average residence time of transcription factors bound to DNA; consequently, highly frequent and short events would appear the same as infrequent and long binding events. To infer these masked kinetics, which are obscured in our mean based ODE model, we use a stochastic binding model to simulate individual molecules of CLOCK:BMAL1 binding to target DNA sites in a well-mixed system (***Gillespie, 1976***). For our simulations, we have used the average number of molecules and effective dissociation rates determined for T28, T32, and T40 hr post-DEX for lung fibroblasts (***Figure 7A***) arising from our previously described ODE model. T28 and T40 represent trough and peak of BMAL1 (***Figure 3D***) respectively, whereas T32 and T40 represent the trough and peak of PER:CRY protein amounts (***Figure 5D***).

Alongside binding of active CLOCK:BMAL1, we also considered target sites bound by CLOCK:BMAL1:CRY1, which are thought to be transcriptionally inactive whilst also blocking target site access to active molecules. At T40, when there is the maximum amount of CLOCK:BMAL1, 95% of the 3,436 target sites would be bound at least once within a minute, changing to ca. 2 minutes at T28 (***Figure 7B***), contributing towards a small degree of heterogeneity. From the perspective of a single promoter at T40 there are ≈3.2 visits per minute by CLOCK:BMAL1, which is reduced down to ≈1.4 visits per minute at T28 (***Figure 7C***), with further reductions in individual cells with lower concentrations of CLOCK:BMAL1 protein (***Figure 7—figure supplement 1***). We then separated the total visits per minute into those occurring as CLOCK:BMAL1 compared to those occurring as the CLOCK:BMAL1:CRY1 complex, finding the latter to remain relatively persistent across time points and making up ≈15% of total visits, mirroring results for our ODE model. Our stochastic model therefore predicts that oscillating amounts of BMAL1 and CRY1 protein amounts, as well as the changing interaction affinity, may actually help preserve the concentration of CLOCK:BMAL1:CRY1 target binding events across circadian time (***Figure 7D***).

Repression of CLOCK:BMAL1 activity by CRY1 requires continuous interaction and hence is limited by concentration. We hypothesised that this would be different for the PER2:CRY1-mediated displacement of CLOCK:BMAL1 from target DNA sites. To test this, we first investigated how the number of visits per minute would be affected by clamping the input values of k_{OFF} and protein to different time points across different observed nuclear volumes. From this, we found that both concentration of protein and k_{OFF} make a substantial contribution to the number of target site binding events (***Figure 7—figure supplement 1***). We then separated the contribution of changing amounts of CLOCK:BMAL1 protein and PER2:CRY1 mediated displacement to visits per minute by calculating the impact of removal of PER2. We find that changing BMAL1 protein abundance accounts for the most variation in number of target site visitations, changing from 1.3 visits at the nadir to 2.7 visits at peak BMAL1 protein (***Figure 7E***). CLOCK:BMAL1 mobility is supported by the action of PER2:CRY1 across all time points, accounting for a maximum 15% of visits at the trough of PER2 protein levels (T40). To explore this relationship further, we tested the impact of altering the levels of PER2 in the stochastic model, choosing four PER2 concentrations, ranging from absent to greater than observed physiological levels (0, 10, 20, and 50 nm) (***Figure 7F***). In the complete absence of PER2, BMAL1 mobility is hampered so that it visits less than three sites per minute. When PER2 spans the physiological range and beyond, a strong relationship in the visits per minute is forecast, rising by a third and in the opposite relationship to site occupancy. Our modelling demonstrates dual modes of action of PER2:CRY1, repression via displacement of CLOCK:BMAL1 from target sites and facilitation of CLOCK:BMAL1 mobility to promote new target site binding (***Figure 7G***). In this sense, PER2 acts both as part of a transcriptional repressor complex and as a facilitator of CLOCK:BMAL1 mobility to bind new target sites (***Cao et al., 2021***).

Discussion

The circadian molecular circuit responds to and modulates an extraordinary number of biological processes, broadly imparted through DNA binding of CLOCK:BMAL1 to E-box sites (***Koike et al., 2012***). Through live cell microscopy of fluorescent ectopically and endogenously expressed circadian

proteins we have sought to understand how the autonomous molecular clock regulates CLOCK:BMAL1 binding to DNA.

Protein abundance and stoichiometry of the circadian circuit

Mathematical modelling demonstrates that low molecular abundances, as observed for core circadian components, lead to rapid internal and cell-to-cell desynchrony, which may be compensated for by strict control of stoichiometries and interactions (***Gonze and Goldbeter, 2006***). In the first instance, protein concentrations of both activators and repressors exert significant influence on amplitude as well as robustness of daily DNA-binding cycles. We found approximately 16,000 BMAL1 and 8000 CRY1 proteins per nucleus, consistent with our earlier reports for PER2 which found 12,000 proteins per nucleus in skin fibroblasts (***Smyllie et al., 2016***). Interestingly, a recent study by Gabriel et al. found approximately an eight-fold difference between CRY1 and PER2 using the U20S, osteocarinoma cancer cell line, highlighting how different cell types and cell lines may diverge and influence the circadian network (***Gabriel et al., 2021***). Similarly, we observed significant disparities of endogenous BMAL1 across a range of cell types, with fibroblasts exhibiting a > 2 fold increase in BMAL1 concentration when compared with chondrocytes (***Figure 3—figure supplement 1***). The impact of cell type variation in protein concentrations and stoichiometries is difficult to discern but may confer tissue-specific sensitives to clock control of output genes without the need for additional regulatory components, or could compensate the system, as evidenced by similar single site visitations despite a fourfold decrease in nuclear volume (***Figure 7—figure supplement 1C***).

Balance in affinity between BMAL1 and CLOCK may facilitate crosstalk

CLOCK was found to be cytoplasmic when ectopically expressed without BMAL1, with nuclear localisation restored upon addition of BMAL1. This suggests BMAL1 oscillations could affect availability of nuclear CLOCK, consistent with several studies (***Kondratov et al., 2003***; ***Kwon et al., 2006***). Our measures of total BMAL1 and CLOCK reveal a concentration ratio of 2:1, possibly reflecting differences in turnover rate, import and export of these two proteins. Strikingly, only 10% of BMAL1 was bound in complex with CLOCK. This ratio of 2:1 is compatible with recent modelling studies defining stoichiometric relationships within the nucleus (***Lee et al., 2011***; ***Kim and Forger, 2012***). Low availability of heterodimeric transcription factor for DNA interactions, when compared with free protein, severely limits the potential to bind DNA, yet this is consistent with allowing other interactions to occur, including those reported with Hypoxia-inducible factor (HIF) and Aryl hydrocarbon receptor (AhR) (***Bagnall et al., 2014***; ***Jaeger and Tischkau, 2016***). Balancing availability of monomeric BMAL1 and CLOCK may therefore enable crosstalk with other pathways, or modulate interactions that have different affinities for monomeric versus heterodimeric CLOCK:BMAL1, as has been reported for CRY1 (***Michael et al., 2017***; ***Xu et al., 2015***).

Impact of cycling CRY1 concentration, binding affinities and mode of repression on the clock

Substantive evidence for direct repression of BMAL1 transactivation by CRY1 now exists (***Gustafson et al., 2017***; ***Xu et al., 2015***). Here, we have shown in live cells that this interaction is not only rhythmic but remarkably strong, with a higher affinity than any other protein pairings we have measured. This strong repression of CLOCK:BMAL1 by CRY1 balances against its low abundance. When acting without PER2, CRY1 exhibits near-persistent repression over 24 hr, likely owing to its regulated interaction with CLOCK and BMAL1, as evidenced by modelling the effect of removal of either cycling BMAL1, CRY1 or binding affinity between the two (***Figure 3F***, ***Figure 5—figure supplement 1F***). This cycle in affinity provides evidence that the mammalian circadian clock also relies on oscillations in the ability of key proteins to heterodimerise one another. The exact mechanisms underlying this regulation of affinity are yet to be determined but could be hypothesised to be an outcome of dimerisation with another partner that hinders or aids binding to CLOCK:BMAL1, such as PER2, or post-translational modifications leading to changes in affinity with CLOCK:BMAL1 (***Ye et al., 2011***; ***Fribourgh et al., 2020***; ***Schmalen et al., 2014***). A ≈25% shift in the diffusion of CRY1 equating to a change in mass close to that of, and in phase with the peak of, PER2 hints at the former proposition but further study is required (***Figures 2G and 3E***).

Individual genes exhibit a range of residence times

We found an average short residence time of 3 s for CLOCK:BMAL1, similar to other DNA binding transcription factors including GR, p53, p65, and STAT1 (***Hettich and Gebhardt, 2018***), potentially optimised to reduce gene expression noise (***Azpeitia and Wagner, 2020***). Here we modelled CLOCK:BMAL1 binding to a number of sites using an average off rate resulting in all sites behaving the same and demonstrating how DNA binding is globally regulated, in contrast with evidence from ChiP-seq, whereby different sites are differentially bound (***Koike et al., 2012***). Presumably, robustly detected peaks found by ChIP-seq represent genes with a slow unbinding rate, such as the E-box sites found in the DBP gene, which is supported by previous live cell imaging characterising a longer 8-s residence time for BMAL1 on a DBP E-box concatemer (***Stratmann et al., 2012***). Altering the unbinding rates leads to a non-linear scaling in the occupation frequency (***Figure 5—figure supplement 1B***), highlighting the importance of regulating this parameter through post-translational modifications such as via phosphorylation of the CLOCK:BMAL1 complex as reported by ***Qin et al., 2015***. Residence time may be tuned individually for different genes to ensure optimal transactivation, especially when considering recruitment of critical co-factors which do not interact with CLOCK:BMAL1 outside of DNA, as the probability of co-occupation increases with binding time. Ultimately, a considerable temporal gulf exists between the elaboration of a circadian rhythm (days) with the time-scale of DNA binding (seconds), altered by the accumulation of protein (hours). Daily changes in BMAL1 protein are moderate, remaining as high as 10,000 molecules per nucleus even at the nadir of expression, resulting in many non-transcriptionally productive interactions of CLOCK:BMAL1 with DNA throughout the circadian cycle; these interactions however may be important, contributing to pioneer factor activity and allowing others genes to activate at a different phase to BMAL1 protein levels (***Klemz et al., 2021***; ***Menet et al., 2014***).

Compromise between E-box visitations and occupancy via PER:CRY mediated displacement

Whereas CRY1 inhibits BMAL1 transactivation via binding and blocking productive interactions with transcriptional coactivators, PER:CRY complexes permit an alternative mode of repression (***Cao et al., 2021***; ***Xu et al., 2015***). We demonstrate that increasing PER:CRY leads to an overall reduction in the ability for CLOCK:BMAL1 to remain bound through direct dimerisation and manipulation of DNA unbinding. Work by Cao and Wang et al revealed PER2 removes CLOCK:BMAL1 in a CRY-dependent manner from E-Boxes via recruitment of CK1 and subsequent phosphorylation of CLOCK, effectively reducing affinity for DNA (***Cao et al., 2021***). Displacive repression of this kind reduces residency time on DNA sites and thus the number of sites bound at any one time. However, reducing residency time increases the rate at which a limited pool of transcription factors can move onto new sites, hence increasing the likelihood of any one gene to be bound and reducing possible cell-to-cell variation. Site-specific residence times, most likely due to cofactor recruitment or chromatin modifications, coupled with this phenomenon would permit some gene targets to exhibit maximal CLOCK:BMAL1 binding beyond the time of the global peak. This supports findings by Menet and colleagues, who highlight groups of genes that have maximal binding events, as determined via ChIP-seq, outside of the zenith of total genome CLOCK:BMAL1 binding (***Menet et al., 2014***). Furthermore, evidence of CLOCK:BMAL1 behaving as a so-called 'kamikaze' transcription factor, a factor most transcriptionally potent when phosphorylated and marked for degradation, implies that in addition to an increase in visitations per minute, transcriptional potency is also upregulated (***Stratmann et al., 2012***). Therefore, despite the relatively high efficiency of CLOCK:BMAL1 binding to DNA, it may spend much of its life performing transcriptionally non-productive tasks until modified via complexes such as PER:CRY. PER:CRY displacement played a significant role, even at its nadir of expression, contributing to 15% of visitations per minutes at the height of DNA binding and CLOCK:BMAL1 concentration (T40). Thus, PER:CRY plays a hidden role of enhancing the mobility of CLOCK:BMAL1 to new DNA sites (***Figure 7G***).

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Genetic reagent (*M. musclus*)	C57BL/6 Venus::BMAL1	***Yang et al., 2020***		Venus sequence inserted before BMAL1 start codon.
Genetic reagent (*M. musclus*)	C57BL/6 Cry1::mRuby3	This paper		CRY1 stop codon replaced with mRuby3
Genetic reagent (*M. musclus*)	C57BL/6 Venus::BMAL1 x CRY1::mRuby3	This paper		Crossed from Venus::BMAL1 and CRY1::mRuby3 mice
Cell line (*M. musculus*)	NIH/3T3	ATCC	CRL-1658	
Transfected construct (*M. musculus*)	pLNT-NLS::EGFP	Vector Builder VB900119-0501njq		Lentiviral construct to express nuclear EGFP.
Transfected construct (*M. musculus*)	pLNT-BMAL1::EGFP or pLNT-BMAL1::RFP	This paper	NCBI reference: NM_007489.4	Lentiviral construct to express fluorescent BMAL1.
Transfected construct (*M. musculus*)	pLNT-BMAL1-L95E::EGFP	This paper	NCBI reference: NM_007489.4	Lentiviral construct to express fluorescent BMAL1 L95E mutant.
Transfected construct (*M. musculus*)	pLNT-BMAL1-V435R::EGFP	This paper	NCBI reference: NM_007489.4	Lentiviral construct to express fluorescent BMAL1 V435R mutant.
Transfected construct (*M. musculus*)	pLNT-EGFP::CLOCK or pLNT-RFP::CLOCK	This paper	NCBI reference: NM_007715.6	Lentiviral construct to express fluorescent CLOCK.
Transfected construct (*M. musculus*)	pLNT-EGFP::PER2	This paper	NCBI reference: NM_011066	Lentiviral construct to express fluorescent PER2.
Transfected construct (*M. musculus*)	pLNT-CRY1::RFP	This paper	NCBI reference: NM_007771.3	Lentiviral construct to express fluorescent CRY1.
Chemical compound, drug	Dexamethasone	Sigma Aldrich	D4902	
Software, algorithm	GraphPad Prism	GraphPad Prism	Version 9	
Software, algorithm	FCCS analysis pipeline	This paper	https://github.com/LoudonLab/FcsAnalysisPipeline, (copy archived at swh:1:rev:b12e9007ed7f8a033485e57c8605e27c67df74f1; ***Koch, 2021***)	

Plasmids

A set of lentivirus transfer plasmids encoding fluorescent fusions of circadian proteins were generated utilising the gateway cloning system as previously described ***Bagnall et al., 2015***. In brief, an initial entry vector was cloned, containing murine coding sequences for: *Bmal1* (NM_007489.4), *Clock* (NM_007715.6), *Cry1* (NM_007771.3), and *Per2* (NM_011066.3). These vectors were then recombined with a target destination vector containing a fluorescent protein sequence to generate a terminal lentivirus vector, in which expression is regulated from the constitutive ubiquitin ligase C promoter. The NLS::EGFP, BMAL1 L95E, and BMAL1 V435R encoding plasmids were all purchased from VectorBuilder.

Primary cell isolates and cell lines

Fibroblasts were isolated from lungs of adult mice. Lung tissue was dissected and homogenised before collagenase-1A (1.5 $\mathrm{mg\,ml^{-1}}$, Cat no. C2674) treatment for 2 hr. The cell suspension was then

filtered using 40 m cell strainers before plating into DMEM (Cat no. D6429) supplemented with 10% fetal bovine serum (HyClone), penicillin-streptomycin (10 U/ml) and amphotericin B (2.5 $g\,ml^{-1}$). Media was refreshed every 2–3 days for 1 week before sub-culturing or experimentation. Cells were sub-cultured for a maximum of 4 passages. SCN slice cultures were prepared as previously described (***Smyllie et al., 2016***) and imaged after 2–3 days after preparation for confocal imaging, or kept for 7 days in culture prior to bioluminescence recording. Cultures derived from separate mice were used as biological replicates.

NIH/3T3 (ATCC CRL-1658) cells were cultured in DMEM supplemented with 10% fetal bovine serum (HyClone).The cells were tested for the absence of mycoplasma using MycoAlert mycoplasma detection kit (Cat. No. LT07-418). Cells were passaged every 2–3 days, maintaining cells till passage 30. Lentivirus transduced derivatives of these cells were made using low passage cultures (6-12). Production of 3rd generation lentivirus and subsequent transduction of NIH/3T3 cells was carried out as previously described (***Bagnall et al., 2015***). Singly transduced cells are referred to with a superscript LV1 prefix before the transgene. Sequential transductions were carried out a minimum of 2 weeks later and derived cells are then termed LV2. Circadian synchronisation of cells was achieved by stimulation with 200 nm dexamethasone (Sigma D4902) for 1 hr before PBS washes and then switched to fresh culture media. Cultures were passaged for biological replicates.

Confocal microscopy

For 2D culture imaging experiments, cells were plated into 35 mm glass bottomed imaging dishes (Greiner Bio-one) at least 6 hr prior to imaging. Measurements were performed using either a ZEISS LSM880 or ZEISS LSM780 microscope equipped with a stage mounted incubator to maintain cells at 37 °C in humidified 5% CO_2; fluorescence image capture was performed using either ZEN 2.1 SP3 FP2 or ZEN 2010b SP1 software, respectively. Fluorescence samples were excited using the most appropriate lasers; making use of an Argon-Ion laser to produce 488 nm or 514 nm excitation or diode laser to produce 561 nm excitation. The appropriate emitted fluorescence spectra were then collected using Quasar GasP array detectors. All images were made using a FLUAR 40 x NA 1.3 oil immersion objective. Nuclear volume recordings were made by collecting a z-stack of images at nyquist rate using a one airy unit pinhole diameter and then analysing images using Imaris (version 7.4). Time-lapse imaging of SCN: fluorescence timelapse recordings of CRY1::mRuby3 in SCN organotypic slices were acquired using Zeiss LSM780/880 inverted confocal system (Zeiss), and maintained at 37 °C. Samples were placed in air-tight glass-bottom dishes (Mattek). Images were acquired using 10 x objective, 30 s scan time per frame, 2 frames per hour, for 6 days for longer time lapse or 60–70 hr for shorter time lapse.

Real-time population bioluminescence recordings

Lung fibroblasts were plated into 35 mm plastic tissue culture dishes (Corning). Cell media was replaced with a HEPES buffered and phenol free DMEM. Additionally, D-luciferin was supplemented into the media 4–24 hr prior to recordings. To prevent gas-exchange, dishes were sealed with grease applied around the edges of the coverslips. Bioluminescence was then recorded by photomultiplier tubes (PMTs; Hamatasu) housed in an enclosed incubator at 37 °C and without CO_2 as described previously (***Loudon et al., 2007***).

Single-molecule fluorescence in situ hybridisation

Clock and *Bmal1* mRNA were visualised using custom probes designed against *Clock* and *Bmal1* murine coding sequences via the Stellaris FISH Probe Designer (Biosearch Technologies Inc). *Clock* and *Bmal1* probes were labeled with the Quasar-570 and Quasar-670 dyes, respectively. Samples were imaged with a wide-field DeltaVision microscope as previously described and spot counting was performed with FISH-quant (***Bagnall et al., 2018***; ***Mueller et al., 2013***).

Fluorescence recovery after photo-bleaching

FRAP was performed by time-lapse imaging of cells prior to and after photobleaching to visualise fluorescence recovery. Photobleaching of EGFP and Venus signals was performed using 488 nm or 514 nm laser lines respectively using circular regions of 5 m diameter (approximately 10% nuclear area) and wholly within the nuclei of cells. Images were recorded every 0.262 s for up to 60 s.

We have used FRAP to infer DNA unbinding rates for BMAL1, see ***Figure 1—figure supplement 2***. The principle of the approach assumes that a combination of diffusion and binding to an unseen immobile substrate affects the speed in which fluorescent proteins move into the recovery area. Different trajectories of recovery therefore inform how the balance between binding and diffusion contributes to the apparent diffusion of the observed molecule. This approach has been utilised many times to characterise the binding times of transcription factors to DNA, including GR, STAT1, p53, and p65 and has been additionally cross validated against single molecule imaging (***Groeneweg et al., 2014***). For our data, the recovery curves of BMAL1::EGFP (co-expressed alongside tagRFP::CLOCK) remained consistent when bleaching different sized nuclear regions indicating that binding contributes to the recovery profile rather than a solely diffusion-led process (***Figure 1—figure supplement 2***). For all subsequent measurements, a circular bleach region was used that was kept consistent across cells and accounted for approximately 10% of nuclear area. FRAP was performed and analysed using the appropriate ZEN software version with recovery curves from the bleached region normalised to total cell fluorescence as well as background fluorescence (empty spaces away from cells). The normalised recovery curve of fluorescence within the bleached region over time, t, was then fit with a reaction binding model

$$f(t) = I_E - I_1 \mathrm{e}^{-\frac{t}{\tau}}, \tag{1}$$

where τ is the residence time (reciprocal of unbinding rate k_{OFF}), I_E and I_1 is the immobile and mobile fraction, respectively (***Sprague and McNally, 2005***).

Fluorescence correlation spectroscopy

Experimental setup

FCS measurements were performed in each cell nucleus using acquisition times of 20 s and a collection volume of 1 airy unit (approximately 0.722 fl and 1.10 fl for 488 nm and 561 nm excitation respectively when using the FLUAR 40 x NA 1.3 oil immersion objective) calibrated in the x-y plane for maximum signal intensity. The effective confocal volumes were calculated via the equation

$$V_{\mathrm{eff}} = (2\pi)^{\frac{3}{2}} w_{xy}^2 w_z = (2\pi)^{\frac{3}{2}} \left(\frac{0.61\lambda}{\mathrm{NA}}\right)^2 \left(\frac{2n\lambda}{\mathrm{NA}^2}\right), \tag{2}$$

where w_{xy} and w_z is the beam width in the xy plane and z axis respectively, with NA as the numerical aperture ($\mathrm{NA} = 1.3$ for our 40 x objective), λ the wavelength of exciting laser and n the refractive index of the immersion oil ($n = 1.515$ in all experiments). The appropriate spectra were collected for each different fluorophore. Laser power was reduced to minimise photo-bleaching whilst maintaining counts per molecule greater than 0.3 kHz.

Fitting

Auto-correlation curves extracted from the Zeiss.fcs files were fit over two rounds using a program written in Python 3; first a global parameter fit executed using a genetic algorithm *differential evolution* (SciPy [***Virtanen et al., 2020***]) generating initial guesses within reasonable parameter bounds was performed, followed by a final stage of non-linear least-squares regression implemented via the *curve fit* (SciPy) package with an arctan loss function. The non-linear regression was regularised using the standard deviation following the calculations by ***Saffarian and Elson, 2003*** which incorporates systematic sources of error at short and long lag times due to the multi-tau correlation algorithm used to compute the correlation curve; at short lag times the averaging introduces uncertainty whilst at the long lag times less data points exist to correlate due to the finite time over which the experiment was run. Poor fits arising from samples expressing only auto-fluorescence or no fluroescence from focus shifts can result in highly non-plausible parameter measurements (diffusion >100 um/s) and concentrations (> 1000 nM) which were removed by robust regression and outlier removal (ROUT) (***Motulsky and Brown, 2006***).

Model selection

The Akaike Information Criterion (AIC) (***Akaike, 1974***) was used to score and select the best fit model with the lowest score, defined as

$$\mathrm{AIC} = 2k - 2\ln(\hat{L}), \tag{3}$$

where k is the number of fitted parameters and $\hat{L}$ the maximum likelihood, equal to the sum of squared errors when using non-linear least squares regression to fit the curves. Results of the model selection for all FCS data sets performed in this study can be found in ***Figure 2—figure supplement 1***.

Interactions: fluorescence cross-correlation spectroscopy

Care was taken for fluorescence cross-correlation spectroscopy (FCCS) measurements to avoid the green channel signal spilling up into the red channel causing false cross-correlation by reducing the laser power and observing the far-red part of the second channel. Control measurements were performed by selectively turning off either 488 or 561 nm lasers and tuning the red channel spectra until there was no cross-correlation due to spill-over. We analysed both sets of auto-correlation and cross-correlation curves from the same measurement and used the same procedure as ***Sadaie et al., 2014*** to calculate the disassociation constant K_D. We again used non-linear least squares regression upon this data, fitting the function

$$\frac{[\mathrm{Complex}]}{[\mathrm{A::GFP}]_{\mathrm{TOTAL}}} = \frac{[\mathrm{B::RFP}]_{\mathrm{TOTAL}} - [\mathrm{Complex}]}{K_D + [\mathrm{B::RFP}]_{\mathrm{TOTAL}} - [\mathrm{Complex}]}, \tag{4}$$

where $[\mathrm{A::GFP}]_{\mathrm{TOTAL}}$ is the total concentration of the protein A fused to a green or yellow fluorescent protein, $[\mathrm{B::RFP}]_{\mathrm{TOTAL}}$ is the total of protein B fused to a red fluorophore and $[\mathrm{Complex}]$ is the concentration of the dimer of A and B proteins. The standard deviation upon K_D was also provided by this algorithm.

Maturation correction

Fluorescent proteins may take minutes or hours to fold correctly before becoming visible, with the invisible fraction becoming substantial if the degradation rate of the protein is comparable to the maturation rate, hence leading to misreports in protein number as measured by FCS. The red fluorescent protein, mRuby3 is known to have a long maturation time of 2.28 (***Balleza et al., 2018***) and CRY1 to have a half-life of approximately 2.1 (***Yoo et al., 2013***), therefore we applied a scaling correction to CRY1::mRuby3 FCS concentration data. To account for the unseen portion, we model the protein in two states; an invisible state, P, and a mature visible fraction, M. Assuming a constant rate of production, k_p, for the immature protein, a maturation rate for the fluorophore of k_m, and a degradation rate for both protein states of k_d we get the set of ordinary differential equations

$$\begin{aligned} \frac{dP}{dt} &= k_p - k_d P - k_m P, \\ \frac{dM}{dt} &= k_m P - k_d M. \end{aligned} \tag{5}$$

These equations may be solved analytically using an integrating factor assuming zero of both protein states at $t = 0$ and so long as the rate constants k_m and k_d are greater than zero. The unknown production rate, k_p, is divided out when computing the ratio of both states by M and taking the limit of the solution as $t \longrightarrow \infty$ to yield the correction factor

$$c = \lim_{t\to\infty}\left(\frac{P(t)+M(t)}{M(t)}\right) = \frac{k_d+k_m}{k_m} = \frac{\tau_m}{\tau_d} + 1, \tag{6}$$

where τ_m and τ_d are the doubling-time and half-life of the maturation and degradation respectively. Using ***Equation (6)***, the half-life for CRY1 and the maturation time of mRuby3 we find a multiplicative factor of $c = 2.083$, which may multiply the observed protein to yield the total concentration of CRY1::mRuby3.

Diffusion rate as a function of mass

When considering normal diffusion due to Brownian motion the diffusion rate, D, may be modelled using the Stokes–Einstein equation (***Einstein, 1905***)

$$D = \frac{k_B T}{8\pi\eta r}, \tag{7}$$

where k_B is the Boltzmann constant, T the temperature in kelvin, η the dynamic viscosity, and r as the radius of the diffusing molecule. Assuming a constant density of spatially equally distributed constituent amino acids, the mass of the molecule grows like r^3 and hence the diffusion rate will be related to the mass of the molecule by

$$D \propto m^{-1/3}, \tag{8}$$

hence a halving in mass will equate to an approximate increase of 1.26 times the diffusion rate.

In vitro binding assays

Expression and purification of recombinant proteins

Biotin Acceptor Peptide (BAP)-tagged CLOCK PAS-AB (mouse CLOCK residues 93–395) was expressed as a His_6-NusA-XL-tagged protein in *Escherichia coli* (*E. coli*) Rosetta2 (DE3) cells. The *E. coli* biotin ligase BirA was expressed as a GST-tagged protein in BL21 (DE3) cells. Protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at an OD_{600} of approximately 0.8 and grown for an additional 16 hr at 18 °C. Cells were centrifuged at 4 °C at 3200 x g, resuspended in 50 mM Tris pH 7.5, 300 mM NaCl, 5% (vol/vol) glycerol, and 5 mM-mercaptoethanol (BME) and lysed using a microfluidizer followed by brief sonication on ice. After clarifying lysate by centrifugation at 4 °C at 140,500 x g for 1 hr, proteins were captured using Ni-NTA resin (Qiagen) or Glutathione Sepharose 4B resin (GE Life Sciences). After extensive washing in in 50 mM Tris pH 7.5, 300 mM NaCl, 5% (vol/vol) glycerol, and 5 mmM BME, the affinity and solubility tags (e.g. His_6-NusA-XL or GST) were cleaved on resin using GST-TEV or His6-TEV protease at 4 °C overnight. Cleaved proteins were collected from the flow-through; GST-BirA was further purified using size exclusion chromatography (SEC) on a Superdex75 column (GE Healthcare) in 50 mM Tris, pH 8.0, 300 mM NaCl, 1 mM dithiothreitol (DTT), and 5% (vol/vol) glycerol, while CLOCK PAS-AB was further purified using SEC in 20 mM HEPES pH 7.5, 125 mM NaCl, 5% (vol/vol) glycerol, and 2 mM Tris(2-carboxyethyl)phosphine (TCEP).

BMAL1 PAS-AB (mouse BMAL1 residues 136–441) was expressed in Sf9 suspension insect cells (Expression systems) as a GST-tagged protein using the baculovirus expression system. Sf9 cells were infected with P3 virus at 1.2×10^6 cells per milliliter and grown for 72 hr at 27 °C before harvesting. Cells were resuspended in resuspension buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 5% (vol/vol) glycerol, and 5 mM-mercaptoethanol (BME)). Cells were lysed using a microfluidizer followed by brief sonication on ice. After clarifying lysate by centrifugation at 140,500 x g for 1 hr at 4 °C, the lysate was bound in batch-mode to Glutathione Sepharose 4B resin (GE Healthcare), washed in resuspension buffer and eluted with 50 mM HEPES pH 7.5, 150 mM NaCl, 5% (vol/vol) glycerol, 5 mM BME, and 25 mM reduced glutathione. The protein was desalted into 50 mM HEPES pH 7, 150 mm NaCl, 5% (vol/vol) glycerol, and 5 mM BME using a HiTrap Desalting column (GE Healthcare) and incubated with GST-TEV protease overnight at 4 °C. The cleaved GST-tag and GST-tagged TEV protease were removed by Glutathione Sepharose 4B chromatography (GE Healthcare) and the BMAL1 PAS-AB was further purified by SEC on a Superdex75 column (GE Healthcare) in 20 mm HEPES pH 7.5, 125 mM NaCl, 5% (vol/vol) glycerol, and 2 mM TCEP. For long-term storage, small aliquots of proteins were frozen in liquid nitrogen and stored at –70 °C.

Biotinylation of CLOCK PAS-AB. For the biotinylation reaction, 100 m BAP-CLOCK PAS-AB was incubated in 20 mM HEPES pH 7.5, 125 mM NaCl, 5% (vol/vol) glycerol, and 2 mM TCEP with 2 mM ATP, 1 m GST-BirA, and 150 m biotin at 4 °C overnight. GST-BirA was removed after the reaction using Glutathione Sepharose 4B resin (GE Healthcare) and excess biotin was separated from the labeled protein by SEC on a Superdex75 column in 20 mm HEPES pH 7.5, 125 mM NaCl, 5% (vol/vol) glycerol, and 2 mM Tris(2-carboxyethyl)phosphine (TCEP). Biotinylation of CLOCK PAS-AB was essentially complete, as determined by incubating the protein with excess streptavidin and resolving complexes on SDS-PAGE. For long-term storage, small aliquots of the biotinylated protein were frozen in liquid nitrogen and stored at –70 °C.

Surface plasmon resonance binding assays

Kinetic binding experiments were conducted on a Biacore X100 +instrument (GE Healthcare) capturing biotinylated CLOCK PAS-AB on a streptavidin-coated SA sensor chip at 100–150 Response Units (RUs). Serial dilutions of BMAL1 PAS-AB from 0.25 to 10 nm were injected in phosphate buffered saline (PBS) over 250 s and dissociated into buffer over 250 s to determine binding kinetics.

Sensorgram data were globally fit to a 1:1 biomolecular binding model with Biacore Evaluation software X100 +version 2.0.1 (GE Healthcare) to determine k_{ON}, k_{OFF} and K_D. χ^2 values lt_1 and $R_{max} \leq 100$ were established as quality cutoffs for acceptable data. See ***Figure 2—figure supplement 2F*** for surface plasmon resonance results between BMAL1 PAS-AB and CLOCK PAS-AB domains.

Animal lines

A previously established Venus::BMAL1 mouse line was used (***Yang et al., 2020***). Additionally, two more mouse lines were generated which included CRY1::mRuby3 (***Figure 3—figure supplement 1***) and a subsequent cross with mice expressing Venus::BMAL1 and PER2::LUC (***Bagnall et al., 2015***). The CRY1::mRuby3 mice were made using a CRISPR-mediated genomic editing approach to introduce a fluorescent sequence via homology-directed repair. Details of methodology and validation of animals can be found in the supplementary materials. Eight- to 10-week-old mice were housed in individual cages in light-tight cabinets (Tecniplast), equipped with activity mouse wheel cages (Actimetrics). Activity was recorded by ClockLab data collection software in 6-min bins (Actimetrics). The mice were maintained at LD cycles (light on at 7 am; light off at 7 pm) for 2 weeks. Activity profiles were generated using ClockLab (Actimetrics) and used to apply Non-Parametric Circadian Rhythm Analysis (NPCRA) to 10 circadian days of wheel-running data, as described previously (***Reppert and Weaver, 2002***), to calculate: Intra-daily Variability (IV): Non-parametric frequency of activity-rest transitions within a day, with a range of between 0 and 2 (e.g. a Sine wave would have a value of 0 and Gaussian noise would have a value of 2). Inter-daily Stability (IS): Matching of activity patterns on day-to-day basis, ranging from 0 (Gaussian noise) to 1 (high-stability). Robust behavioural activity is characterised by low IV and high IS. ClockLab (Actimetrics) was used to generate double-plotted actograms with onsets of activity and phase angle of entrainment was calculated from 10 days of wheel-running data measuring the difference in time of the point in the entraining cycle (lights on) against the onset of activity.

Generation of CRY1::mRuby3 mouse line

We used CRISPR-Cas9 to generate C terminally tagged alleles for Cry1, see ***Figure 3—figure supplement 2***. Two sgRNA targeting the STOP codon of the gene were selected using the Sanger WTSI website (***Hodgkins et al., 2015***) that adhered to our criteria for off target predictions (guides with mismatch (MM) of 0, 1 or 2 for elsewhere in the genome were discounted, and MM3 were tolerated if predicted off targets were not exonic). sgRNA sequences, wih PAM site indicated in italics, (aactgata cggtaaatactt-AGG and cggcagagcagtaactgata-CGG) were purchased as crRNA oligos, which were annealed with tracrRNA (both oligos supplied by IDT) in sterile, RNase free injection buffer (TrisHCl 1 mM, pH 7.5, EDTA 0.1 mM) by combining 2.5 mg crRNA with 5 mg tracrRNA and heating to 95°C, which was allowed to slowly cool to room temperature.

For our donor repair template, we used the EASI-CRISPR long-ssDNA strategy (***Quadros et al., 2017***), which comprised of the mRuby3 gene with linker flanked by 132 and 143 nt homology arms synthesised by a Biotinylation PCR and on-column denaturation method (***Bennett et al., 2021***; ***Figure 3—figure supplement 3A***). For embryo microinjection, the annealed sgRNA was complexed with Cas9 protein (New England Biolabs) at room temperature for 10 min, before addition of long ssDNA donor (final concentrations; sgRNA 20 ng/ml, Cas9 protein 20 ng/ml, lssDNA 10 ng/ml). CRISPR reagents were directly microinjected into C57BL6/J (Envigo) zygote pronuclei using standard protocols. Zygotes were cultured overnight and the resulting two-cell embryos surgically implanted into the oviduct of day 0.5 post-coitum pseudopregnant mice. Potential founder mice were screened by PCR (***Figure 3—figure supplement 2A***), using primers that flank the sgRNA sites (Cut test F taca ctatgctcacggggac and Cut test R accacgtcctcttcagaacc), which both identifies editing activity in the form of InDels from NHEJ repair, and can also detect larger products indicating HDR (***Figure 3—figure supplement 2A***). Pups 18, 19, and 22, which gave positive products in PCR reactions, were sequenced by amplifying again with the cut test F/R primers using high fidelity Phusion polymerase (NEB), gel extracted and subcloned into pCRblunt (Invitrogen) and Sanger sequenced with M13 Forward and Reverse primers. All pups showed perfect sequence integration and were bred with a WT C57BL6/J to confirm germline transmission.

Mathematical modelling

Modelling aims and assumptions

We sought to model how the core circadian transcription factor, CLOCK:BMAL1, binds to specific E-BOX DNA sites over daily cycles in concentration and interactions with the key repressors CRYPTOCHROME1 (CRY1) and PERIOD2 (PER2). We have opted to neglect explicitly modelling the non-specific DNA interactions, such as sliding, hopping and intersegmental transfer, as we have no direct measurements of these properties. Instead we chose to allow the specific site on rate (k_{ON}) to represent all protein-DNA processes required to achieve binding to an specific site by fitting k_{ON} alongside other ON rates. CRY1 and PER2 repress the activity of BMAL1 through direct binding of the transactivation domain (TAD) to block transcriptional potential and the promotion of weaker binding to DNA, respectively. We have assumed that PER2 may interact with BMAL1 and CLOCK:BMAL1 only via CRY1 with the same affinity that CRY1 alone has for BMAL1. To constrain the on rates during fitting, we have used the measured disassociation constants, K_D, between CLOCK-BMAL1 (***Figure 2G***), CLOCK:BMAL1-EBOX (***Huang et al., 2012***), CRY1-PER2 (***Figure 4C***), and the rhythmic CRY1-BMAL1 K_D (***Figure 3F***). All protein-protein and protein-DNA interactions are modelled as explicit dimerisation events leading to a new species dependent on an ON and OFF rate. A summary of the parameters, K_D values and which parameters were proposed during fitting is given in ***Table 1***. Following the convention when defining chemical reactions, square brackets are used to signify concentrations of the species within. To aid understanding of the reactions being modelled we describe the species participating in reactions as familiar initialisations, for example [CB] represents the concentration of the CLOCK:BMAL1 heterodimer and [C1] for CRY1. Consequently, further dimerisations or bound states are denoted by concatenations of these initialisations, for example [CBC1P2] for the CLOCK:BMAL1:CRY1:PER2 tetramer or [CBS] for CLOCK:BMAL1 bound to a specific DNA (S) site.

Dimerisation

Hetero-dimerisation of two species $[A]$ and $[B]$ proceeds to the dimer $[AB]$ via the reaction

$$[A] + [B] \underset{k_{\mathrm{OFF}}}{\overset{k_{\mathrm{ON}}}{\rightleftharpoons}} [AB], \tag{9}$$

where where k_{ON} ($\mathrm{nm}^{-1}\,\mathrm{s}^{-1}$) and k_{OFF} (s^{-1}) are the forward and backwards rates respectively (***Sadaie et al., 2014***), often referred to as the association and dissociation rate constants. In equilibrium, the forward rate of reaction is equal to the backward rate resulting in the definition of the disassociation constant

$$K_D = \frac{[A][B]}{[AB]}, \tag{10}$$

defined in terms of the ON and OFF rates as

$$K_D = \frac{k_{\mathrm{OFF}}}{k_{\mathrm{ON}}}. \tag{11}$$

A stronger interaction is represented as a smaller K_D value as the rate to the disassociated is smaller than the association rate. In the limit of long times, $t \to \infty$, the concentration of the dimer $[AB]$ in equilibrium becomes

$$[AB]_{\mathrm{eq}} = \frac{[A]_0 + [B]_0 + K_D - \sqrt{([A]_0 + [B]_0 + K_d)^2 - 4[A]_0[B]_0}}{2}, \tag{12}$$

where a subscript 0 denotes the initial concentration. Alternatively, assuming no production or degradation terms exist we may simulate analytically intractable multiple interactions by simulating a coupled ODE model until equilibrium concentrations are reached. For all ODE modelling, we defined equilibrium as less than a 1% deviation in molecular concentrations over the last 20% of simulated time points. In all cases equilibrium was established in less than 30 min of simulated time, smaller than the window over which experimental FCS time point measurements were performed.

Ordinary differential equation model of DNA binding

Systems of ordinary differential equations (ODE), modelling the concentrations of molecular species, were solved in the Python three programming language to reflect measured interactions between

different molecules and DNA. All interactions are modelled as explicit dimerisation events yielding a new molecular species. ODEs were solved in time as an initial value problem using the LSODA solver as implemented in the SciPy *odeint* function (**Virtanen et al., 2020**) and ran until equilibrium concentrations were reached, typically reached in less than 30 min of simulated time. The system of ODEs are

$$\frac{d[CB]}{dt} = -d_{\mathrm{ON}}[CB][C1P2]+d_{\mathrm{OFF}}[CBC1P2]-d_{\mathrm{ON}}[CB][C1]+d_{\mathrm{OFF}}[CBC1]-k_{\mathrm{ON}}[CB][S] +k_{\mathrm{OFF}}[CBS]+ b_{\mathrm{ON}}[C][B] - b_{\mathrm{OFF}}[CB],$$

$$\frac{d[S]}{dt} = -k_{\mathrm{ON}}[CB][S] + k_{\mathrm{OFF}}[CBS] - k_{\mathrm{ON}}[CBC1P2][S] + R_{\mathrm{OFF}}[CBC1P2S] - k_{\mathrm{ON}}[CBC1][S] + k_{\mathrm{OFF}}[CBC1S],$$

$$\frac{d[CBS]}{dt} = k_{\mathrm{ON}}[CB][S] - k_{\mathrm{OFF}}[CBS] - d_{\mathrm{ON}}[CBS][C1] + d_{\mathrm{OFF}}[CBC1S] - d_{\mathrm{ON}}[CBS][C1P2] + d_{\mathrm{OFF}}[CBC1P2S],$$

$$\frac{d[C1]}{dt} = -d_{\mathrm{ON}}[CB][C1]+d_{\mathrm{OFF}}[CBC1]-d_{\mathrm{ON}}[CBS][C1]+d_{\mathrm{OFF}}[CBC1S]-a_{\mathrm{ON}}[C1][P2] +a_{\mathrm{OFF}}[C1P2]- d_{\mathrm{ON}}[B][C1] + d_{\mathrm{OFF}}[BC1],$$

$$\frac{d[CBC1]}{dt} = -a_{\mathrm{ON}}[CBC1][P2] + a_{\mathrm{OFF}}[CBC1P2] + d_{\mathrm{ON}}[CB][C1] - d_{\mathrm{OFF}}[CBC1] - k_{\mathrm{ON}}[CBC1][S] + k_{\mathrm{OFF}}[CBC1S] + b_{\mathrm{ON}}[C][BC1] - b_{\mathrm{OFF}}[CBC1],$$

$$\frac{d[CBC1S]}{dt} = d_{\mathrm{ON}}[CBS][C1] - d_{\mathrm{OFF}}[CBC1S] - a_{\mathrm{ON}}[CBC1S][P2] + a_{\mathrm{OFF}}[CBC1P2S] + k_{\mathrm{ON}}[CBC1][S] - k_{\mathrm{OFF}}[CBC1S],$$

$$\frac{d[P2]}{dt} = -a_{\mathrm{ON}}[CBC1][P2] + a_{\mathrm{OFF}}[CBC1P2] - a_{\mathrm{ON}}[C1][P2] + a_{\mathrm{OFF}}[C1P2] - a_{\mathrm{ON}}[CBC1S][P2] + a_{\mathrm{OFF}}[CBC1P2S] - a_{\mathrm{ON}}[BC1][P2] + a_{\mathrm{OFF}}[BC1P2],$$

$$\frac{d[C1P2]}{dt} = -d_{\mathrm{ON}}[CB][C1P2] + d_{\mathrm{OFF}}[CBC1P2] + a_{\mathrm{ON}}[C1][P2] - a_{\mathrm{OFF}}[C1P2] - d_{\mathrm{ON}}[CBS][C1P2] + d_{\mathrm{OFF}}[CBC1P2S] - d_{\mathrm{ON}}[B][C1P2] + d_{\mathrm{OFF}}[BC1P2],$$

$$\frac{d[CBC1P2S]}{dt} = k_{\mathrm{ON}}[CBC1P2][S] - R_{\mathrm{OFF}}[CBC1P2S] + a_{\mathrm{ON}}[CBC1S][P2] - a_{\mathrm{OFF}}[CBC1P2S] + d_{\mathrm{ON}}[CBS][C1P2] - d_{\mathrm{OFF}}[CBC1P2S],$$

$$\frac{d[CBC1P2]}{dt} = a_{\mathrm{ON}}[CBC1][P2] - a_{\mathrm{OFF}}[CBC1P2] + d_{\mathrm{ON}}[CB][C1P2] - d_{\mathrm{OFF}}[CBC1P2] - k_{\mathrm{ON}}[CBC1P2][S] + R_{\mathrm{OFF}}[CBC1P2S] + b_{\mathrm{ON}}[C][BC1P2] - b_{\mathrm{OFF}}[CBC1P2],$$

$$\frac{d[C]}{dt} = -b_{\mathrm{ON}}[C][B] + b_{\mathrm{OFF}}[CB] - b_{\mathrm{ON}}[C][BC1] + b_{\mathrm{OFF}}[CBC1] - b_{\mathrm{ON}}[C][BC1P2] + b_{\mathrm{OFF}}[CBC1P2],$$

$$\frac{d[B]}{dt} = -b_{\mathrm{ON}}[C][B] + b_{\mathrm{OFF}}[CB] - d_{\mathrm{ON}}[B][C1] + d_{\mathrm{OFF}}[BC1] - d_{\mathrm{ON}}[B][C1P2] + d_{\mathrm{OFF}}[BC1P2],$$

$$\frac{d[BC1]}{dt} = d_{\mathrm{ON}}[B][C1] - d_{\mathrm{OFF}}[BC1] - b_{\mathrm{ON}}[C][BC1] + b_{\mathrm{OFF}}[CBC1] - a_{\mathrm{ON}}BC1[P2] + a_{\mathrm{OFF}}[BC1P2],$$

$$\frac{d[BC1P2]}{dt} = d_{\mathrm{ON}}[B][C1P2] - d_{\mathrm{OFF}}[BC1P2] - b_{\mathrm{ON}}[C][BC1P2] + b_{\mathrm{OFF}}[CBC1P2] + a_{\mathrm{ON}}[BC1][P2] - a_{\mathrm{OFF}}[BC1P2].$$

A genetic algorithm, implemented in *differential evolution* (SciPy [**Virtanen et al., 2020**]), was utilised to fit the unknown parameters in the ODE model via Chi-squared minimisation to experimental k_{OFF} mean and standard error on the mean using an in silico value, $\bar{k}_{OFF}$, generated by the model. All non-dimerised species concentrations, as measured experimentally, were introduced for each of the seven time-points – a 24 hr time span sampled every 4 hr – into the model as inputs alongside measured disassociation constants to constrain fitted OFF rates as a function of proposed ON rates, reducing the number of fitted parameters. A summary of the parameters in the model is given in **Table 1**. During fitting the in silico k_{OFF} value was calculated by allowing all species to reach equilibrium after setting all DNA bound species to zero following an initial run of the model, the resultant equilibrium concentrations of bound and free molecules were used to calculate the off rate (**Figure 5—figure supplement 1A**). The average apparent DNA unbinding rate $\bar{k}_{OFF}$, which is analogous to the same rate as experimentally measured in FRAP, is simulated following the method by **Röding et al., 2019** through rearranging **Equation (11)** for the off rate

$$\bar{k}_{OFF} = k_{\mathrm{ON}}K_D = k_{\mathrm{ON}}\frac{[\text{Unbound CB}][\text{Unbound Sites}]}{[\text{Bound Sites}]}, \tag{13}$$

with

$$[\text{Unbound CB}] = [CB]_{\text{eq}} + [CBC1]_{\text{eq}} + [CBC1P2]_{\text{eq}}, \tag{14}$$

$$[\text{Unbound Sites}] = [S]_{\text{eq}}, \tag{15}$$

$$[\text{Bound sites}] = [CBS]_{\text{eq}} + [CBC1S]_{\text{eq}} + [CBC1P2S]_{\text{eq}} \tag{16}$$

where eq denotes concentrations at equilibrium as $t \to \infty$. The apparent $\bar{k}_{OFF}$ is an average of the CLOCK:BMAL1-DNA binding OFF rate k_{OFF} and CLOCK:BMAL1:CRY1:PER2-DNA OFF rate R_{OFF} weighted by their respective relative concentrations, with increasing levels of CRY1:PER2 increasing $\bar{k}_{OFF}$ as $k_{\text{OFF}} < R_{\text{OFF}}$. The fitted parameters are given in ***Table 1*** and predicted in silico $\bar{k}_{OFF}$ values for WT and PER2 KO can be seen in (***Figure 5—figure supplement 1C***). Knocking out PER2 (keeping all other species and parameters the same as wild type values) removes all rhythmic regulation of $\bar{k}_{OFF}$ and ensures that CLOCK:BMAL1 is bound for longer at all time points such that the number of bound specific sites (S) also increases for all time points (***Figure 5—figure supplement 1E***). Locking BMAL1, CRY1, PER2 and the interaction between BMAL1 and CRY1 to their mean value between 24 and 48 hr post-dexamethosone (DEX) treatment, reveals that setting BMAL1 to its mean value significantly alters both bound and free from DNA CLOCK:BMAL1:CRY1 whilst locking the other rhythmic components has little impact (***Figure 5—figure supplement 1F***).

Upper and lower bounds on one-at-a-time (OAT) sensitivity analysis (***Figure 6A–B***) were generated by running the fitted model for both an estimate of the smallest and largest number of target sites, 1000 and 10,000 respectively, with the mean representing the mean number of target sites from ChIP data, namely 3,436. We may estimate the number of target sites for CLOCK:BMAL1 from previous studies investigating high confidence sites that BMAL1 binds to in ChIP-seq, with ***Table 2*** outlining the reports and peaks measured via ChIP-seq that were used in estimating the number of target sites used in our mathematical modelling. In addition to the OAT analyses in ***Figure 6A–B*** we also examined how changing amounts of CRY1:PER2 alters the residence time of CLOCK:BMAL1 on DNA, demonstrating how CRY1:PER2 readily promotes removal from DNA in a non-linear fashion over a physiologically plausible range of concentrations (***Figure 5—figure supplement 1B***).

Stochastic DNA binding model

Stochastic binding simulations in Python three utilised the Gillespie algorithm (***Gillespie, 2002***) through the StochPy library (***Maarleveld et al., 2013***) to simulate a reduced topology, considering

Table 3. Stochastic model reactions and propensities.
Counter for arrivals by CLOCK:BMAL1 (CB) without CRY1 ($C1$) to previously unbound sites S converting them to S_0 given by A_{CB} as well as counters for marked site M binding represented by B_X, and unbinding, U_X, by species X. The size of the system is given by $\Omega = 1 \times 10^{-9} N_A V$, where V is the volume in liters and is used to convert ON rate quantities with dimensions nm^{-1}s^{-1} into particle^{-1} s^{-1}. k_{ON} is the same value as previously fitted for the ODE model given in ***Table 1***.

No.	Reaction	Propensity
1	$CB + S \longrightarrow CBS + A_{CB}$	$(k_{\text{ON}}/\Omega) \cdot CB \cdot S$
2	$CB + S_0 \longrightarrow CBS$	$(k_{\text{ON}}/\Omega) \cdot CB \cdot S_0$
3	$CBS \longrightarrow CB + S_0$	$k_{\text{OFF}} \cdot CBS$
4	$CBC1 + S \longrightarrow CBC1S$	$(k_{\text{ON}}/\Omega) \cdot CBC1 \cdot S$
5	$CBC1S \longrightarrow CBC1 + S$	$k_{\text{OFF}} \cdot CBC1S$
6	$CBC1 + S_0 \longrightarrow CBC1S_0$	$(k_{\text{ON}}/\Omega) \cdot CBC1 \cdot S_0$
7	$CBC1S_0 \longrightarrow CBC1 + S_0$	$k_{\text{OFF}} \cdot CBC1S_0$
8	$CB + M \longrightarrow CBM + B_{CB}$	$(k_{\text{ON}}/\Omega) \cdot CB \cdot M$
9	$CBM \longrightarrow CB + M + U_{CB}$	$k_{\text{OFF}} \cdot CBM$
10	$CBC1 + M \longrightarrow CBC1M + B_{CBC1}$	$(k_{\text{ON}}/\Omega) \cdot CBC1 \cdot M$
11	$CBC1M \longrightarrow CBC1 + M + U_{CBC1}$	$k_{\text{OFF}} \cdot CBC1M$

only CLOCK:BMAL1 and CLOCK:BMAL1:CRY1 binding to sites with the addition of 1 extra marked site, M, and using the fitted ON/OFF rates from the ODE model. The ***Table 3*** gives the reactions and propensities that are modelled.

Thirty runs over 60 min were used to generate mean and standard deviations with times to reach 95% of all sites at least once determined via fitting of an inverse exponential to the number of unique site visits counted via the variable A_{CB}. The time for available CLOCK:BMAL1 complexes to bind 95% of all binding sites at least once is calculated by fitting the recovery curve $f(t) = 1 - \exp(-\lambda t)$ to normalised stochastic trajectories of $S_{tot} - S$ ($S_{tot} = 3436$), see ***Figure 7—figure supplement 1B***, and then converting the recovery rate λ using the equation

$$\tau_{95\%} = \frac{\ln(20)}{\lambda}. \tag{17}$$

Visits per minute to a single site were calculated by counting binding and unbinding to M, which possesses the same ON and OFF rates as other target-sites. Assessment of the contribution of PER2 mediated displacement was performed by setting PER2 concentration to zero (KO) in the ODE model and using the simulated OFF rate in a parallel run to wild-type (WT) runs (***Figure 5—figure supplement 1B***), with the reduced number of visits attributed to the slower OFF rate. Furthermore, we observed the same behaviour in this reduced stochastic model, when compared to the full ODE model, for PER2 KO as the mean and standard deviation of the number of sites bound by CLOCK:BMAL1 in both WT and KO conditions, ***Figure 7—figure supplement 1A***, being comparable to the ODE model results in (***Figure 5—figure supplement 1E***). Finally, to assess the differences that would be induced by different nuclear volumes, as seen between different cell types, we ran the stochastic model at the same molecular concentrations over two volumes; a small volume of 240 fl representative of a typical mouse embryonic fibroblast (MEF) or various immune cell types (see ***Figure 3—figure supplement 1E***) and 926 fl as measured for our lung fibroblasts used throughout this study, ***Figure 7—figure supplement 1C***. We note little difference in the rate at which CLOCK:BMAL1 visits the single marked site M, indicating that the increase in DNA sites comparatively to the number of molecules at a smaller nuclear volume was balanced by the increase in ON rate due to the now higher concentration of DNA.

Acknowledgements

Supported by the Biotechnology and Biological Sciences Research Council, UK (awards BB/P017347/1 and BB/P017355/1 to ASIL and MHH), MRC core funding (MC_U105170643 to MHH), the National Institutes of Health (award GM107069 and GM141849 to CLP) and a University of California Office of the President Chancellor's Postdoctoral Fellowship (to JLF). ASIL is a Wellcome Investigator (107851/Z/15/Z) and AAK is supported by a Wellcome 4 year PhD studentship (216416/Z/19/Z).

Additional information

Funding

Funder	Grant reference number	Author
Biotechnology and Biological Sciences Research Council	BB/P017347/1	James S Bagnall Nicola Begley Andrew SI Loudon
Biotechnology and Biological Sciences Research Council	BB/P017355/1	Nicola J Smyllie Michael H Hastings
Medical Research Council	MC_U105170643	Michael H Hastings
National Institutes of Health	GM107069	Carrie L Partch
National Institutes of Health	GM141849	Carrie L Partch
Wellcome Trust	107851/Z/15/Z	Andrew SI Loudon

Funder	Grant reference number	Author
Wellcome Trust	216416/Z/19/Z	Alex A Koch
University of California		Jennifer L Fribourgh

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Alex A Koch, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing; James S Bagnall, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing; Nicola J Smyllie, Formal analysis, Investigation, Methodology, Visualization, Writing – review and editing; Nicola Begley, Investigation; Antony D Adamson, Investigation, Resources, Validation; Jennifer L Fribourgh, Formal analysis, Investigation, Visualization; David G Spiller, Investigation, Methodology, Resources, Writing – review and editing; Qing-Jun Meng, Resources, Writing – review and editing; Carrie L Partch, Conceptualization, Formal analysis, Investigation, Resources, Visualization, Writing – review and editing; Korbinian Strimmer, Methodology, Software, Supervision; Thomas A House, Methodology, Software, Supervision, Validation, Writing – review and editing; Michael H Hastings, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing; Andrew SI Loudon, Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review and editing

Author ORCIDs

Alex A Koch http://orcid.org/0000-0002-0592-331X
James S Bagnall http://orcid.org/0000-0003-1735-5855
Nicola J Smyllie http://orcid.org/0000-0002-6324-1421
Nicola Begley http://orcid.org/0000-0003-3601-9152
Antony D Adamson http://orcid.org/0000-0002-5408-0013
David G Spiller http://orcid.org/0000-0003-2502-6787
Qing-Jun Meng http://orcid.org/0000-0002-9426-8336
Carrie L Partch http://orcid.org/0000-0002-4677-2861
Korbinian Strimmer http://orcid.org/0000-0001-7917-2056
Thomas A House http://orcid.org/0000-0001-5835-8062
Michael H Hastings http://orcid.org/0000-0001-8576-6651
Andrew SI Loudon http://orcid.org/0000-0003-3648-445X

Ethics

All experimental procedures were carried out in accordance with the Animals (Scientific Procedures) Act of 1986, UK (Licence number PP7901495).

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.73976.sa1
Author response https://doi.org/10.7554/eLife.73976.sa2

Additional files

Supplementary files

- Transparent reporting form

Data availability

Modelling and analtyical code has been made publicly available via GitHub. The FCS analysis software is at https://github.com/LoudonLab/FcsAnalysisPipeline (copy archived at swh:1:rev:b12e9007ed7f8a033485e57c8605e27c67df74f1), and the modeling link is https://github.com/LoudonLab/CLOCK-BMAL1-DNA-Binding. Source Data files have been provided for all FCS measurements and FRAP measurements in Figures 1, 2, 3 ,4, and 6.

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Koch AA, Bagnall JS	2021	FCS Analysis Pipeline	https://github.com/LoudonLab/FcsAnalysisPipeline	GitHub, FcsAnalysisPipeline
Koch AA	2021	Modelling for Quantification of protein abundance and interaction defines a mechanism for operation of the circadian clock	https://github.com/LoudonLab/CLOCK-BMAL1-DNA-Binding	GitHub, CLOCK-BMAL1-DNA-Binding

References

Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**:716–723. DOI: https://doi.org/10.1109/TAC.1974.1100705

Aryal RP, Kwak PB, Tamayo AG, Gebert M, Chiu PL, Walz T, Weitz CJ. 2017. Macromolecular assemblies of the mammalian circadian clock. *Molecular Cell* **67**:770-782.. DOI: https://doi.org/10.1016/j.molcel.2017.07.017, PMID: 28886335

Azpeitia E, Wagner A. 2020. Short residence times of dna-bound transcription factors can reduce gene expression noise and increase the transmission of information in a gene regulation system. *Frontiers in Molecular Biosciences* **7**:67. DOI: https://doi.org/10.3389/fmolb.2020.00067, PMID: 32411721

Bagnall J, Leedale J, Taylor SE, Spiller DG, White MRH, Sharkey KJ, Bearon RN, Sée V. 2014. Tight control of hypoxia-inducible factor-alpha transient dynamics is essential for cell survival in hypoxia. *The Journal of Biological Chemistry* **289**:5549–5564. DOI: https://doi.org/10.1074/jbc.M113.500405, PMID: 24394419

Bagnall J, Boddington C, Boyd J, Brignall R, Rowe W, Jones NA, Schmidt L, Spiller DG, White MRH, Paszek P. 2015. Quantitative dynamic imaging of immune cell signalling using lentiviral gene transfer. *Integrative Biology* **7**:713–725. DOI: https://doi.org/10.1039/c5ib00067j, PMID: 25990200

Bagnall J, Boddington C, England H, Brignall R, Downton P, Alsoufi Z, Boyd J, Rowe W, Bennett A, Walker C, Adamson A, Patel NMX, O'Cualain R, Schmidt L, Spiller DG, Jackson DA, Müller W, Muldoon M, White MRH, Paszek P. 2018. Quantitative analysis of competitive cytokine signaling predicts tissue thresholds for the propagation of macrophage activation. *Science Signaling* **11**:540. DOI: https://doi.org/10.1126/scisignal.aaf3998, PMID: 30042130

Bajar BT, Wang ES, Lam AJ, Kim BB, Jacobs CL, Howe ES, Davidson MW, Lin MZ, Chu J. 2016. Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and fret reporting. *Scientific Reports* **6**:20889. DOI: https://doi.org/10.1038/srep20889, PMID: 26879144

Balleza E, Kim JM, Cluzel P. 2018. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nature Methods* **15**:47–51. DOI: https://doi.org/10.1038/nmeth.4509, PMID: 29320486

Bennett H, Aguilar-Martinez E, Adamson AD. 2021. Crispr-mediated knock-in in the mouse embryo using long single stranded dna donors synthesised by biotinylated pcr. *Methods (San Diego, Calif.)* **191**:3–14. DOI: https://doi.org/10.1016/j.ymeth.2020.10.012, PMID: 33172594

Beytebiere JR, Trott AJ, Greenwell BJ, Osborne CA, Vitet H, Spence J, Yoo SH, Chen Z, Takahashi JS, Ghaffari N, Menet JS. 2019. Tissue-specific bmal1 cistromes reveal that rhythmic transcription is associated with rhythmic enhancer-enhancer interactions. *Genes & Development* **33**:294–309. DOI: https://doi.org/10.1101/gad.322198.118, PMID: 30804225

Bhadra U, Thakkar N, Das P, Pal Bhadra M. 2017. Evolution of circadian rhythms: from bacteria to human. *Sleep Medicine* **35**:49–61. DOI: https://doi.org/10.1016/j.sleep.2017.04.008, PMID: 28619182

Buhr ED, Takahashi JS. 2013. Molecular components of the mammalian circadian clock. *Handbook of Experimental Pharmacology*. Springer. p. 3–27. DOI: https://doi.org/10.1007/978-3-642-25950-0_1, PMID: 23604473

Cao X, Yang Y, Selby CP, Liu Z, Sancar A. 2021. Molecular mechanism of the repressive phase of the mammalian circadian clock. *PNAS* **118**:e2021174118. DOI: https://doi.org/10.1073/pnas.2021174118, PMID: 33443219

Chiou YY, Yang Y, Rashid N, Ye R, Selby CP, Sancar A. 2016. Mammalian period represses and de-represses transcription by displacing clock-bmal1 from promoters in a cryptochrome-dependent manner. *PNAS* **113**:E6072–E6079. DOI: https://doi.org/10.1073/pnas.1612917113, PMID: 27688755

Dross N, Spriet C, Zwerger M, Müller G, Waldeck W, Langowski J. 2009. Mapping egfp oligomer mobility in living cell nuclei. *PLOS ONE* **4**:e5041. DOI: https://doi.org/10.1371/journal.pone.0005041, PMID: 19347038

Dyar KA, Hubert MJ, Mir AA, Ciciliot S, Lutter D, Greulich F, Quagliarini F, Kleinert M, Fischer K, Eichmann TO, Wright LE, Peña Paz MI, Casarin A, Pertegato V, Romanello V, Albiero M, Mazzucco S, Rizzuto R, Salviati L, Biolo G, et al. 2018. Transcriptional programming of lipid and amino acid metabolism by the skeletal muscle circadian clock. *PLOS Biology* **16**:e2005886. DOI: https://doi.org/10.1371/journal.pbio.2005886, PMID: 30096135

Einstein A. 1905. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen Der Physik* **322**:549–560. DOI: https://doi.org/10.1002/andp.19053220806

Fribourgh JL, Srivastava A, Sandate CR, Michael AK, Hsu PL, Rakers C, Nguyen LT, Torgrimson MR, Parico GCG, Tripathi S, Zheng N, Lander GC, Hirota T, Tama F, Partch CL. 2020. Dynamics at the serine loop underlie differential affinity of cryptochromes for clock:bmal1 to control circadian timing. *eLife* **9**:e55275. DOI: https://doi.org/10.7554/eLife.55275, PMID: 32101164

Fustin JM, O'Neill JS, Hastings MH, Hazlerigg DG, Dardente H. 2009. Cry1 circadian phase in vitro: wrapped up with an e-box. *Journal of Biological Rhythms* **24**:16–24. DOI: https://doi.org/10.1177/0748730408329267, PMID: 19150926

Gabriel CH, Del Olmo M, Zehtabian A, Jäger M, Reischl S, van Dijk H, Ulbricht C, Rakhymzhan A, Korte T, Koller B, Grudziecki A, Maier B, Herrmann A, Niesner R, Zemojtel T, Ewers H, Granada AE, Herzel H, Kramer A. 2021. Live-cell imaging of circadian clock protein dynamics in crispr-generated knock-in cells. *Nature Communications* **12**:3796. DOI: https://doi.org/10.1038/s41467-021-24086-9, PMID: 34145278

Gekakis N, Staknis D, Nguyen HB, Davis FC, Wilsbacher LD, King DP, Takahashi JS, Weitz CJ. 1998. Role of the clock protein in the mammalian circadian mechanism. *Science (New York, N.Y.)* **280**:1564–1569. DOI: https://doi.org/10.1126/science.280.5369.1564, PMID: 9616112

Gibbs J, Ince L, Matthews L, Mei J, Bell T, Yang N, Saer B, Begley N, Poolman T, Pariollaud M, Farrow S, DeMayo F, Hussell T, Worthen GS, Ray D, Loudon A. 2014. An epithelial circadian clock controls pulmonary inflammation and glucocorticoid action. *Nature Medicine* **20**:919–926. DOI: https://doi.org/10.1038/nm.3599, PMID: 25064128

Gillespie DT. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* **22**:403–434. DOI: https://doi.org/10.1016/0021-9991(76)90041-3

Gillespie DT. 2002. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**:2340–2361. DOI: https://doi.org/10.1021/j100540a008

Gonze D, Goldbeter A. 2006. Circadian rhythms and molecular noise. *Chaos (Woodbury, N.Y.)* **16**:026110. DOI: https://doi.org/10.1063/1.2211767, PMID: 16822042

Groeneweg FL, van Royen ME, Fenz S, Keizer VIP, Geverts B, Prins J, de Kloet ER, Houtsmuller AB, Schmidt TS, Schaaf MJM. 2014. Quantitation of glucocorticoid receptor DNA-binding dynamics by single-molecule microscopy and FRAP. *PLOS ONE* **9**:e90532. DOI: https://doi.org/10.1371/journal.pone.0090532, PMID: 24632838

Gustafson CL, Parsley NC, Asimgil H, Lee HW, Ahlbach C, Michael AK, Xu H, Williams OL, Davis TL, Liu AC, Partch CL. 2017. A slow conformational switch in the bmal1 transactivation domain modulates circadian rhythms. *Molecular Cell* **66**:447-457.. DOI: https://doi.org/10.1016/j.molcel.2017.04.011, PMID: 28506462

Hettich J, Gebhardt JCM. 2018. Transcription factor target site search and gene regulation in a background of unspecific binding sites. *Journal of Theoretical Biology* **454**:91–101. DOI: https://doi.org/10.1016/j.jtbi.2018.05.037, PMID: 29870697

Hodgkins A, Farne A, Perera S, Grego T, Parry-Smith DJ, Skarnes WC, Iyer V. 2015. WGE: a CRISPR database for genome engineering. *Bioinformatics (Oxford, England)* **31**:3078–3080. DOI: https://doi.org/10.1093/bioinformatics/btv308, PMID: 25979474

Huang N, Chelliah Y, Shan Y, Taylor CA, Yoo SH, Partch C, Green CB, Zhang H, Takahashi JS. 2012. Crystal structure of the heterodimeric clock:bmal1 transcriptional activator complex. *Science (New York, N.Y.)* **337**:189–194. DOI: https://doi.org/10.1126/science.1222804, PMID: 22653727

Jaeger C, Tischkau SA. 2016. Role of aryl hydrocarbon receptor in circadian clock disruption and metabolic dysfunction. *Environmental Health Insights* **10**:133–141. DOI: https://doi.org/10.4137/EHI.S38343, PMID: 27559298

Kim JK, Forger DB. 2012. A mechanism for robust circadian timekeeping via stoichiometric balance. *Molecular Systems Biology* **8**:630. DOI: https://doi.org/10.1038/msb.2012.62, PMID: 23212247

Klemz S, Wallach T, Korge S, Rosing M, Klemz R, Maier B, Fiorenza NC, Kaymak I, Fritzsche AK, Herzog ED, Stanewsky R, Kramer A. 2021. Protein phosphatase 4 controls circadian clock dynamics by modulating CLOCK/BMAL1 activity. *Genes & Development* **35**:1161–1174. DOI: https://doi.org/10.1101/gad.348622.121, PMID: 34301769

Koch A. 2021. FcsAnalysisPipeline. swh:1:rev:b12e9007ed7f8a033485e57c8605e27c67df74f1. GitHub. https://archive.softwareheritage.org/swh:1:dir:7be2074d3b3bdeae950b9764635cefd44cdcf610;origin=https://github.com/LoudonLab/FcsAnalysisPipeline;visit=swh:1:snp:e84fef6938a998ada7620ab50ac06ca75b49c11f;anchor=swh:1:rev:b12e9007ed7f8a033485e57c8605e27c67df74f1

Koike N, Yoo SH, Huang HC, Kumar V, Lee C, Kim TK, Takahashi JS. 2012. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science (New York, N.Y.)* **338**:349–354. DOI: https://doi.org/10.1126/science.1226339, PMID: 22936566

Kondratov RV, Chernov MV, Kondratova AA, Gorbacheva VY, Gudkov AV, Antoch MP. 2003. Bmal1-dependent circadian oscillation of nuclear clock: posttranslational events induced by dimerization of transcriptional activators of the mammalian clock system. *Genes & Development* **17**:1921–1932. DOI: https://doi.org/10.1101/gad.1099503, PMID: 12897057

Krieger JW, Singh AP, Bag N, Garbe CS, Saunders TE, Langowski J, Wohland T. 2015. Imaging fluorescence (cross-) correlation spectroscopy in live cells and organisms. *Nature Protocols* **10**:1948–1974. DOI: https://doi.org/10.1038/nprot.2015.100, PMID: 26540588

Kwon I, Lee J, Chang SH, Jung NC, Lee BJ, Son GH, Kim K, Lee KH. 2006. Bmal1 shuttling controls transactivation and degradation of the clock/bmal1 heterodimer. *Molecular and Cellular Biology* **26**:7318–7330. DOI: https://doi.org/10.1128/MCB.00337-06, PMID: 16980631

Lee Y, Chen R, Lee H, Lee C. 2011. Stoichiometric relationship among clock proteins determines robustness of circadian rhythms. *The Journal of Biological Chemistry* **286**:7033–7042. DOI: https://doi.org/10.1074/jbc.M110.207217, PMID: 21199878

Liu AC, Tran HG, Zhang EE, Priest AA, Welsh DK, Kay SA. 2008. Redundant function of rev-erbalpha and beta and non-essential role for bmal1 cycling in transcriptional regulation of intracellular circadian rhythms. *PLOS Genetics* **4**:e1000023. DOI: https://doi.org/10.1371/journal.pgen.1000023, PMID: 18454201

Loudon ASI, Meng QJ, Maywood ES, Bechtold DA, Boot-Handford RP, Hastings MH. 2007. The biology of the circadian ck1 tau mutation in mice and syrian hamsters: A tale of two species. *Cold Spring Harbor Symposia on Quantitative Biology* **72**:261–271. DOI: https://doi.org/10.1101/sqb.2007.72.073, PMID: 18522517

Maarleveld TR, Olivier BG, Bruggeman FJ. 2013. Stochpy: A comprehensive, user-friendly tool for simulating stochastic biological processes. *PLOS ONE* **8**:e79345. DOI: https://doi.org/10.1371/journal.pone.0079345, PMID: 24260203

Menet JS, Pescatore S, Rosbash M. 2014. Clock:bmal1 is a pioneer-like transcription factor. *Genes & Development* **28**:8–13. DOI: https://doi.org/10.1101/gad.228536.113, PMID: 24395244

Michael AK, Fribourgh JL, Chelliah Y, Sandate CR, Hura GL, Schneidman-Duhovny D, Tripathi SM, Takahashi JS, Partch CL. 2017. Formation of a repressive complex in the mammalian circadian clock is mediated by the secondary pocket of cry1. *PNAS* **114**:1560–1565. DOI: https://doi.org/10.1073/pnas.1615310114, PMID: 28143926

Motulsky HJ, Brown RE. 2006. Detecting outliers when fitting data with nonlinear regression – a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics* **7**:123. DOI: https://doi.org/10.1186/1471-2105-7-123, PMID: 16526949

Mueller F, Senecal A, Tantale K, Marie-Nelly H, Ly N, Collin O, Basyuk E, Bertrand E, Darzacq X, Zimmer C. 2013. Fish-quant: automatic counting of transcripts in 3d fish images. *Nature Methods* **10**:277–278. DOI: https://doi.org/10.1038/nmeth.2406, PMID: 23538861

Oishi Y, Hayashi S, Isagawa T, Oshima M, Iwama A, Shimba S, Okamura H, Manabe I. 2017. Bmal1 regulates inflammatory responses in macrophages by modulating enhancer rna transcription. *Scientific Reports* **7**:7086. DOI: https://doi.org/10.1038/s41598-017-07100-3, PMID: 28765524

Qin X, Mori T, Zhang Y, Johnson CH. 2015. Per2 differentially regulates clock phosphorylation versus transcription by reciprocal switching of ck1epsilon activity. *Journal of Biological Rhythms* **30**:206–216. DOI: https://doi.org/10.1177/0748730415582127, PMID: 25994100

Quadros RM, Miura H, Harms DW, Akatsuka H, Sato T, Aida T, Redder R, Richardson GP, Inagaki Y, Sakai D, Buckley SM, Seshacharyulu P, Batra SK, Behlke MA, Zeiner SA, Jacobi AM, Izu Y, Thoreson WB, Urness LD, Mansour SL, et al. 2017. Easi-crispr: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssdna donors and crispr ribonucleoproteins. *Genome Biology* **18**:92. DOI: https://doi.org/10.1186/s13059-017-1220-4, PMID: 28511701

Reppert SM, Weaver DR. 2002. Coordination of circadian timing in mammals. *Nature* **418**:935–941. DOI: https://doi.org/10.1038/nature00965, PMID: 12198538

Rey G, Cesbron F, Rougemont J, Reinke H, Brunner M, Naef F. 2011. Genome-wide and phase-specific dna-binding rhythms of bmal1 control circadian output functions in mouse liver. *PLOS Biology* **9**:e1000595. DOI: https://doi.org/10.1371/journal.pbio.1000595, PMID: 21364973

Röding M, Lacroix L, Krona A, Gebäck T, Lorén N. 2019. A highly accurate pixel-based frap model based on spectral-domain numerical methods. *Biophysical Journal* **116**:1348–1361. DOI: https://doi.org/10.1016/j.bpj.2019.02.023, PMID: 30878198

Sadaie W, Harada Y, Matsuda M, Aoki K. 2014. Quantitative in vivo fluorescence cross-correlation analyses highlight the importance of competitive effects in the regulation of protein-protein interactions. *Molecular and Cellular Biology* **34**:3272–3290. DOI: https://doi.org/10.1128/MCB.00087-14, PMID: 24958104

Saffarian S, Elson EL. 2003. Statistical analysis of fluorescence correlation spectroscopy: The standard deviation and bias. *Biophysical Journal* **84**:2030–2042. DOI: https://doi.org/10.1016/S0006-3495(03)75011-5, PMID: 12609905

Schmalen I, Reischl S, Wallach T, Klemz R, Grudziecki A, Prabu JR, Benda C, Kramer A, Wolf E. 2014. Interaction of circadian clock proteins cry1 and per2 is modulated by zinc binding and disulfide bond formation. *Cell* **157**:1203–1215. DOI: https://doi.org/10.1016/j.cell.2014.03.057, PMID: 24855952

Shafer OT, Rosbash M, Truman JW. 2002. Sequential nuclear accumulation of the clock proteins period and timeless in the pacemaker neurons of *Drosophila melanogaster*. *The Journal of Neuroscience* **22**:5946–5954. DOI: https://doi.org/10.1523/JNEUROSCI.22-14-05946.2002, PMID: 12122057

Smyllie NJ, Pilorz V, Boyd J, Meng QJ, Saer B, Chesham JE, Maywood ES, Krogager TP, Spiller DG, Boot-Handford R, White MRH, Hastings MH, Loudon ASI. 2016. Visualizing and quantifying intracellular behavior and abundance of the core circadian clock protein period2. *Current Biology* **26**:1880–1886. DOI: https://doi.org/10.1016/j.cub.2016.05.018, PMID: 27374340

Smyllie NJ, Bagnall J, Koch AA, Niranjan D, Polidarova L, Chesham JE, Chin JW, Partch CL, Loudon ASI, Hastings MH. 2022. Cryptochrome proteins regulate the circadian intracellular behavior and localization of per2 in mouse suprachiasmatic nucleus neurons. *PNAS* **119**:e2113845119. DOI: https://doi.org/10.1073/pnas.2113845119, PMID: 35046033

Sprague BL, McNally JG. 2005. Frap analysis of binding: proper and fitting. *Trends in Cell Biology* **15**:84–91. DOI: https://doi.org/10.1016/j.tcb.2004.12.001, PMID: 15695095

Stratmann M, Suter DM, Molina N, Naef F, Schibler U. 2012. Circadian dbp transcription relies on highly dynamic bmal1-clock interaction with e boxes and requires the proteasome. *Molecular Cell* **48**:277–287. DOI: https://doi.org/10.1016/j.molcel.2012.08.012, PMID: 22981862

Tsekouras K, Siegel AP, Day RN, Pressé S. 2015. Inferring diffusion dynamics from fcs in heterogeneous nuclear environments. *Biophysical Journal* **109**:7–17. DOI: https://doi.org/10.1016/j.bpj.2015.05.035, PMID: 26153697

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, et al. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**:261–272. DOI: https://doi.org/10.1038/s41592-020-0772-5, PMID: 32094914

Wu Y, Tang D, Liu N, Xiong W, Huang H, Li Y, Ma Z, Zhao H, Chen P, Qi X, Zhang EE. 2017. Reciprocal regulation between the circadian clock and hypoxia signaling at the genome level in mammals. *Cell Metabolism* **25**:73–85. DOI: https://doi.org/10.1016/j.cmet.2016.09.009, PMID: 27773697

Xu H, Gustafson CL, Sammons PJ, Khan SK, Parsley NC, Ramanathan C, Lee HW, Liu AC, Partch CL. 2015. Cryptochrome 1 regulates the circadian clock through dynamic interactions with the bmal1 c terminus. *Nature Structural & Molecular Biology* **22**:476–484. DOI: https://doi.org/10.1038/nsmb.3018, PMID: 25961797

Yang N, Smyllie NJ, Morris H, Gonçalves CF, Dudek M, Pathiranage DRJ, Chesham JE, Adamson A, Spiller DG, Zindy E, Bagnall J, Humphreys N, Hoyland J, Loudon ASI, Hastings MH, Meng QJ. 2020. Quantitative live imaging of venus::bmal1 in a mouse model reveals complex dynamics of the master circadian clock regulator. *PLOS Genetics* **16**:e1008729. DOI: https://doi.org/10.1371/journal.pgen.1008729, PMID: 32352975

Ye R, Selby CP, Ozturk N, Annayev Y, Sancar A. 2011. Biochemical analysis of the canonical model for the mammalian circadian clock. *The Journal of Biological Chemistry* **286**:25891–25902. DOI: https://doi.org/10.1074/jbc.M111.254680, PMID: 21613214

Ye R, Selby CP, Chiou YY, Ozkan-Dagliyan I, Gaddameedhi S, Sancar A. 2014. Dual modes of clock:bmal1 inhibition mediated by cryptochrome and period proteins in the mammalian circadian clock. *Genes & Development* **28**:1989–1998. DOI: https://doi.org/10.1101/gad.249417.114, PMID: 25228643

Yoo SH, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, Buhr ED, Siepka SM, Hong HK, Oh WJ, Yoo OJ, Menaker M, Takahashi JS. 2004. Period2::luciferase real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *PNAS* **101**:5339–5346. DOI: https://doi.org/10.1073/pnas.0308709101, PMID: 14963227

Yoo S-H, Mohawk JA, Siepka SM, Shan Y, Huh SK, Hong H-K, Kornblum I, Kumar V, Koike N, Xu M, Nussbaum J, Liu X, Chen Z, Chen ZJ, Green CB, Takahashi JS. 2013. Competing e3 ubiquitin ligases govern circadian periodicity bydegradation of cry in nucleus and cytoplasm. *Cell* **152**:1091–1105. DOI: https://doi.org/10.1016/j.cell.2013.01.055, PMID: 23452855

Yu L, Lei Y, Ma Y, Liu M, Zheng J, Dan D, Gao P. 2021. A comprehensive review of fluorescence correlation spectroscopy. *Frontiers in Physics* **9**:644450. DOI: https://doi.org/10.3389/fphy.2021.644450

RESEARCH ARTICLE

Figures and figure supplements

Quantification of protein abundance and interaction defines a mechanism for operation of the circadian clock

Alex A Koch *et al*

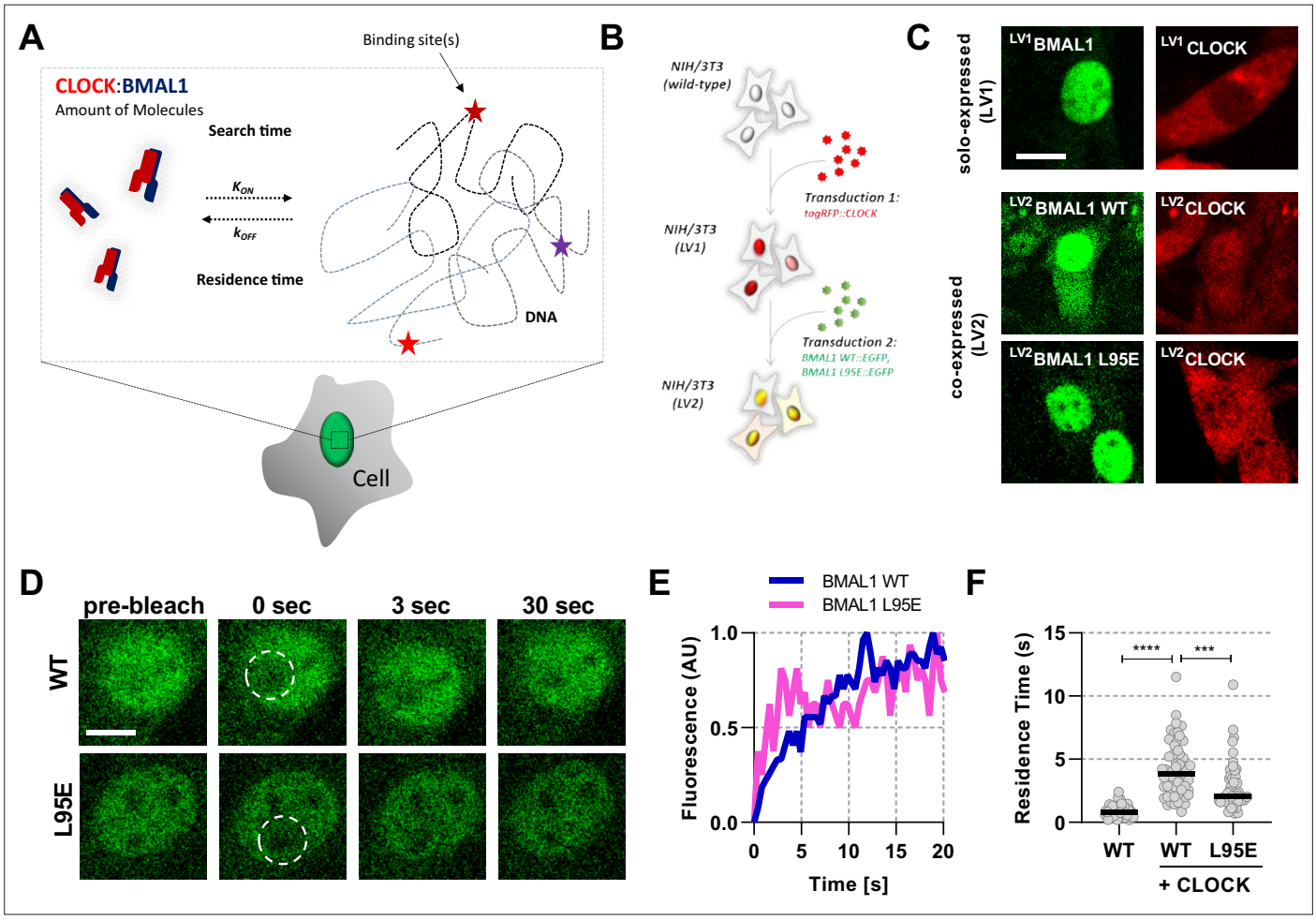

Figure 1. Short-lived DNA binding of BMAL1 and CLOCK. (**A**) Schematic representation of parameters regulating CLOCK:BMAL1 dimers binding to target DNA sites. (**B**) NIH/3T3 cells are either singularly or sequentially transduced to express fluorescent fusions with CLOCK or BMAL1 (wildtype and mutant variants).(**C**) Confocal microscopy images of cells solo-expressing (LV1) either tagRFP::CLOCK or BMAL1::EGFP or co-expressing (LV2) them together (including BMAL1 L95E DNA-binding mutant). (**D**) Confocal microscopy images for photobleaching of ^{LV2}BMAL1::EGFP-RFP::CLOCK labelled cells, either with wild-type or BMAL1 L95E DNA binding mutant. Images show nuclei and highlight region of bleaching. (**E**) Representative fluorescence recovery curves of bleach region for B. following normalisation. (**F**) Residence time calculated as the inverse of kOFF (s^{-1}), determined from fitting the recovery data with a single component binding model (n = 69, 58, and 51 cells). Bar represents median values. Source data for panel F available as ***Figure 1—source data 1***.

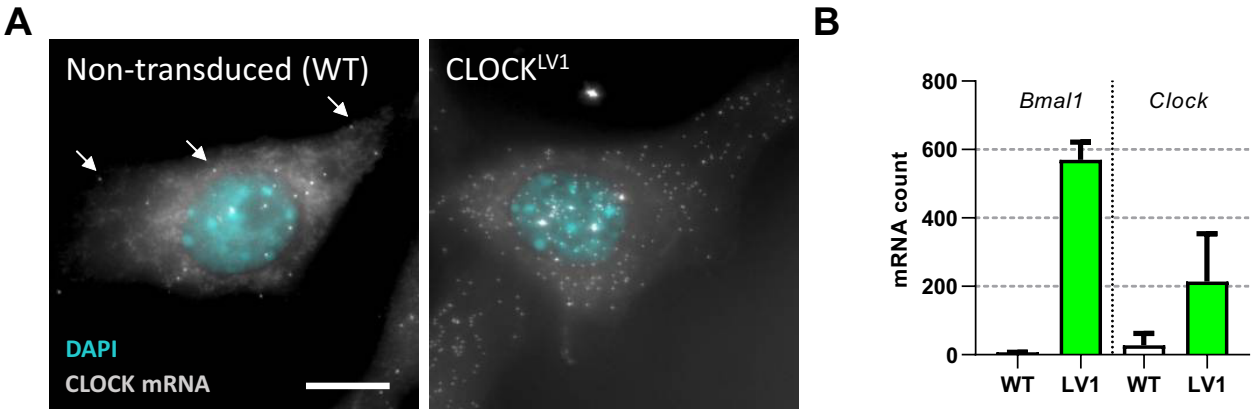

Figure 1—figure supplement 1. Ectopically expressed mRNA is the major form in a lentivirus transduced system. (**A**) NIH/3T3 cells were stained for CLOCK mRNA by single-molecule fluorescent in situ hybridisation (smFISH). (**B**) Mature mRNA was counted from images of many single cells to determine the mRNA content per cell for both non-transduced and cells transduced to express EGFP::CLOCK from a ubiquitin ligase C promoter (n = 2177 cells for WT and n = 155 cells for LV1). Increased nuclear bright dots can be seen for the transduced cells. These bright dots correspond to sites of transcription which is increased beyond two copies due to the multiple sites of integration following transduction. Full details for summary statistics can be found in ***Figure 1—figure supplement 1—source data 1***.

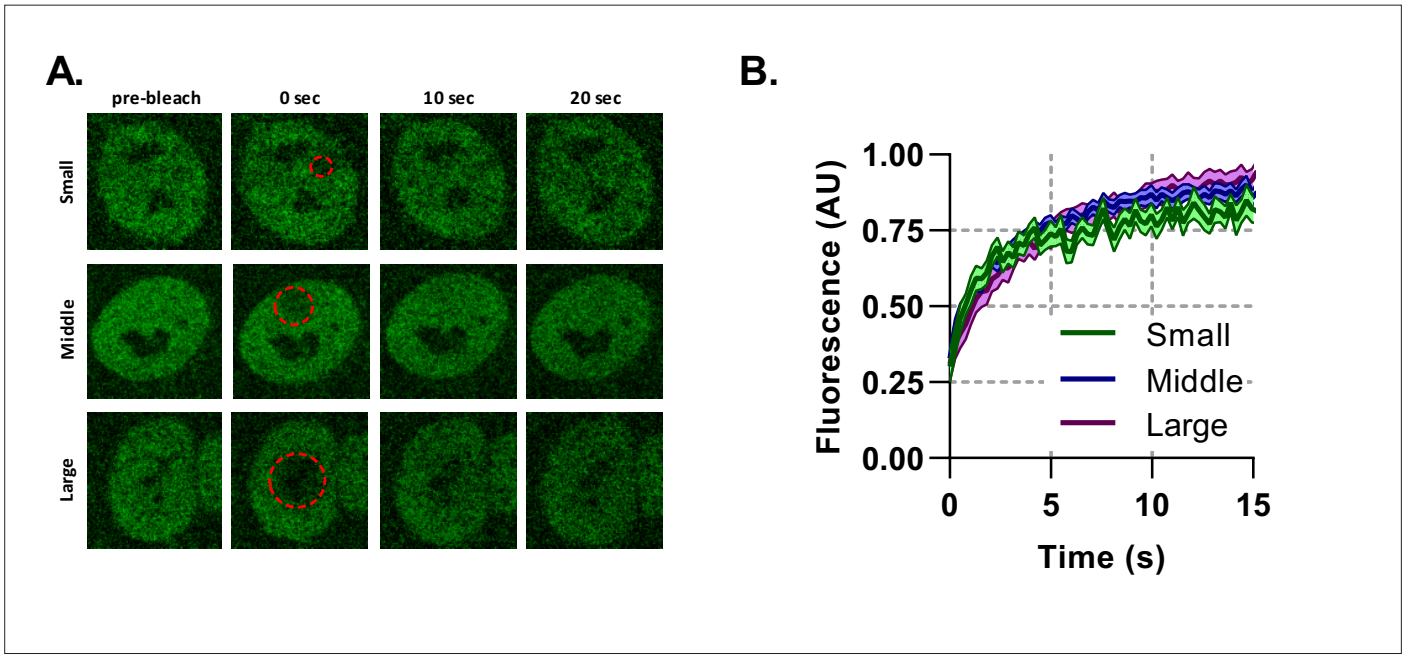

Figure 1—figure supplement 2. Binding plays a significant role in BMAL1 mobility. Fluorescence recovery after photobleaching for NIH/3T3 cells transduced with tagRFP::CLOCK and BMAL1:EGFP. (A) Imaging protocol was performed on BMAL1::EGFP signal. Regions of photobleaching are shown as a red-dotted line which is increased in size. (B) The bleached region recovery curves are shown as averages of all cells with an SEM error envelope (n = 14, 20, and 15).

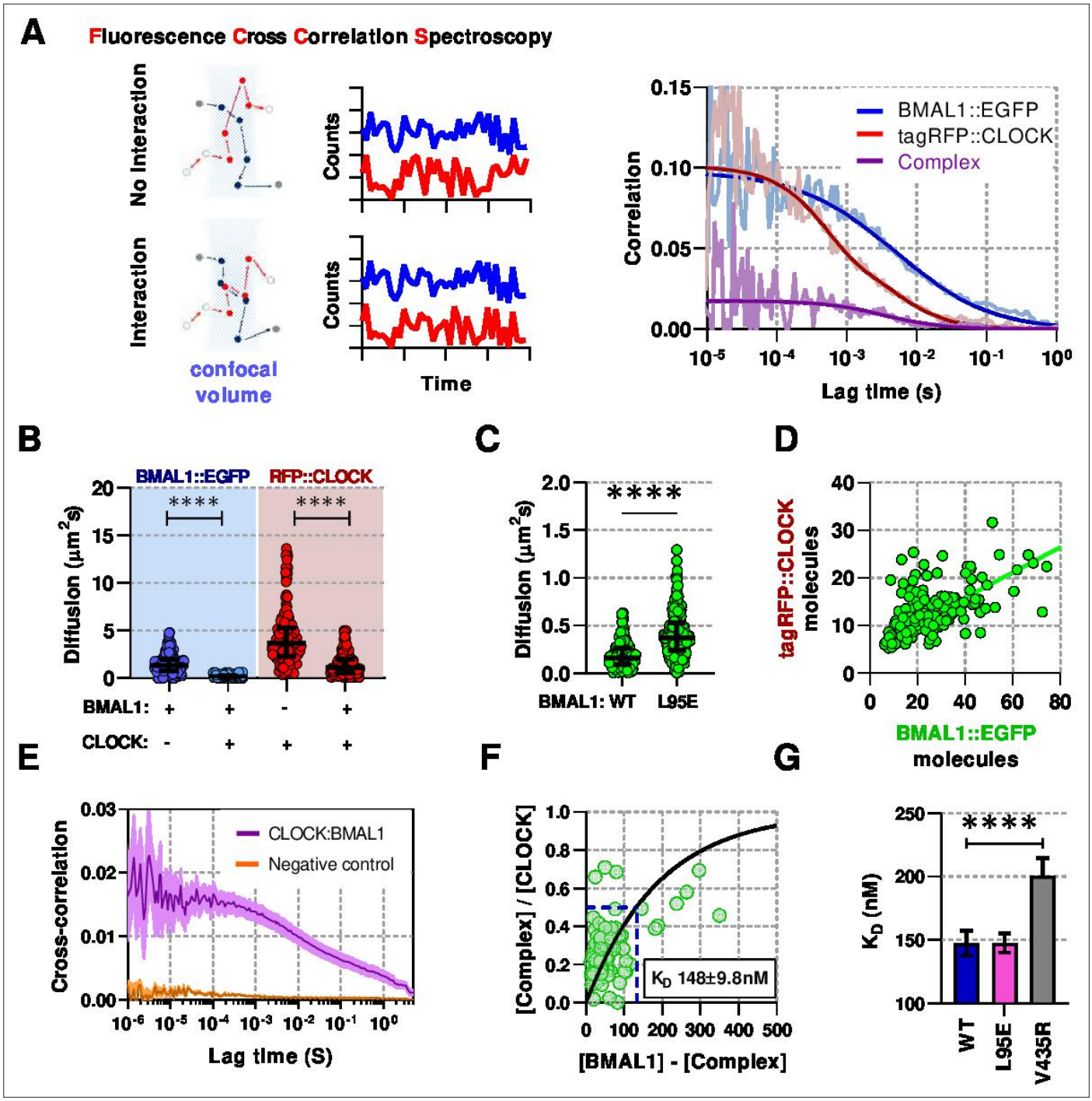

Figure 2. Live-cell interaction measurements demonstrate BMAL1 and CLOCK mobility is regulated by dimerisation and DNA binding. (**A**) Schematic of confocal volume used in FCCS with corresponding photon count traces. Interaction may be seen by correlation between both channels. Representative auto- and cross- correlation data showing raw data and fit lines for monomeric and complexed fluorescent proteins. (**B**) FCS data showing diffusion for BMAL1 and CLOCK in solo- and co-expressed conditions (n = 173, 152, 198, and 185 cells). (**C**) FCS results for BMAL1::EGFP diffusion for NIH/3T3 cells that co-express tagRFP::CLOCK. Data shown is for comparison of BMAL1 as either wild-type of L95E DNA-binding mutant. Bars show median and interquartile range. (**D**) Correlation of nuclear protein quantification showing relationship of BMAL1::EGFP with tagRFP::CLOCK for both wildtype and DNA binding mutant (n = 221 cells from three biological replicates). (**E**) Average cross-correlation curves for BMAL1::EGFP (WT) with tagRFP::CLOCK (n = 140) compared to a non-interacting control of NLS::EGFP co-expressed with tagRFP::CLOCK (n = 408). (**F**) Dissociation plot from FCCS data for BMAL1::WT and tagRFP::CLOCK. (**G**) Summary of calculated dissociation constants across all conditions, including BMAL1 dimerisation mutant, V435R

Figure 2 continued on next page

Figure 2 continued

(n = 156, 274, and 244). Mann-Whitney non-parametric test to determine significance (values are denoted as $p > 0.05$ ns, $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ *** and $p < 0.0001$ ****). Source data for panels B,C available as ***Figure 2—source data 1*** and panel E as ***Figure 2—source data 2***.

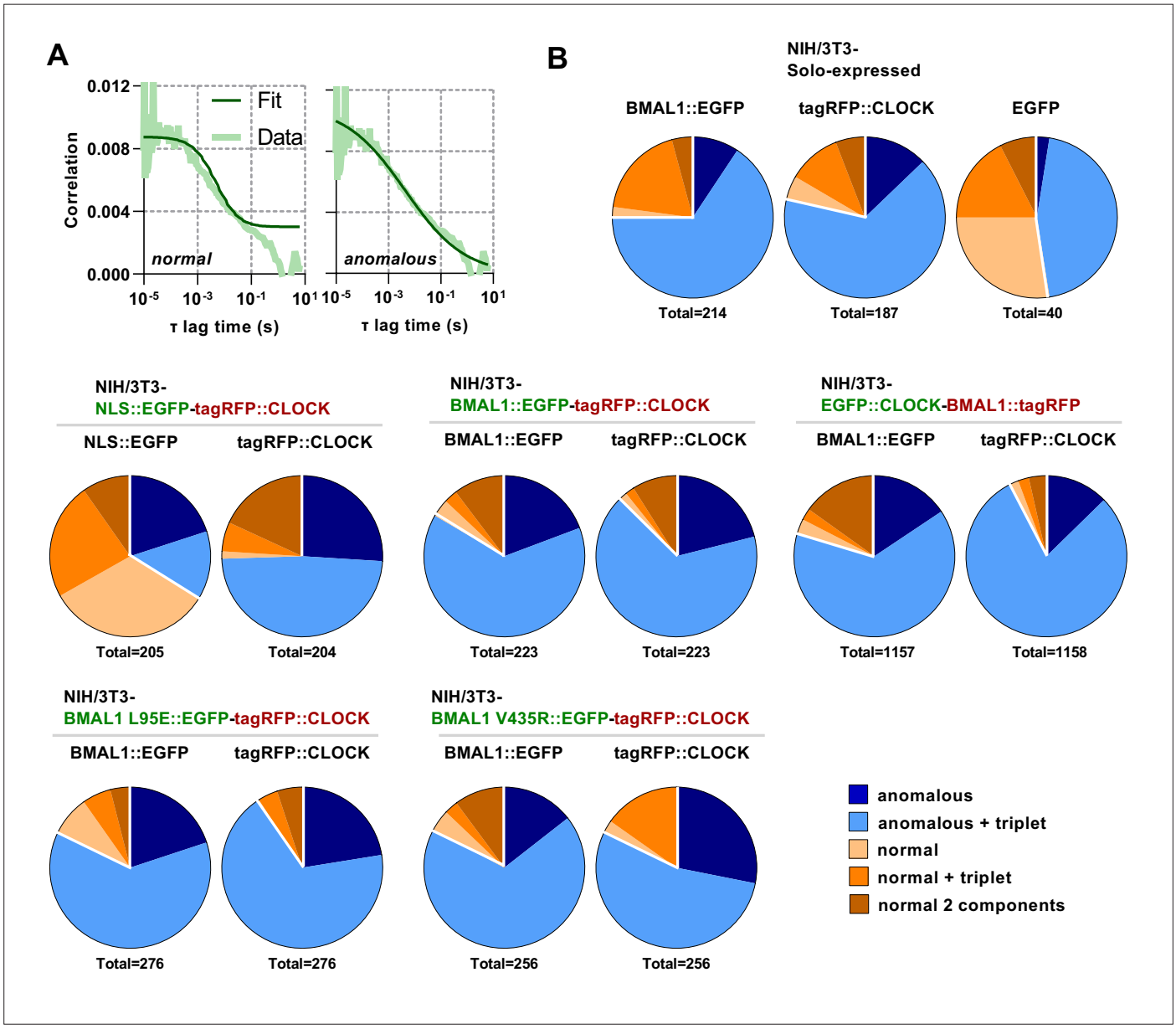

Figure 2—figure supplement 1. Anomalous diffusion best fits protein movement. (**A**) Representative normal and anomalous model fits for BMAL1::EGFP FCS data sets for cells transduced with lentivirus to express BMAL1, CLOCK or control fluorescent proteins. (**B**) Summary data sets for all five models considered. Each model was fit to each measurement and the fit with the lowest AIC score selected. A single pie chart was generated for cells carrying a single fluorescent label whereas two pie charts are plotted for cells with multiple labels, corresponding to the analysed green or red fluorescence.

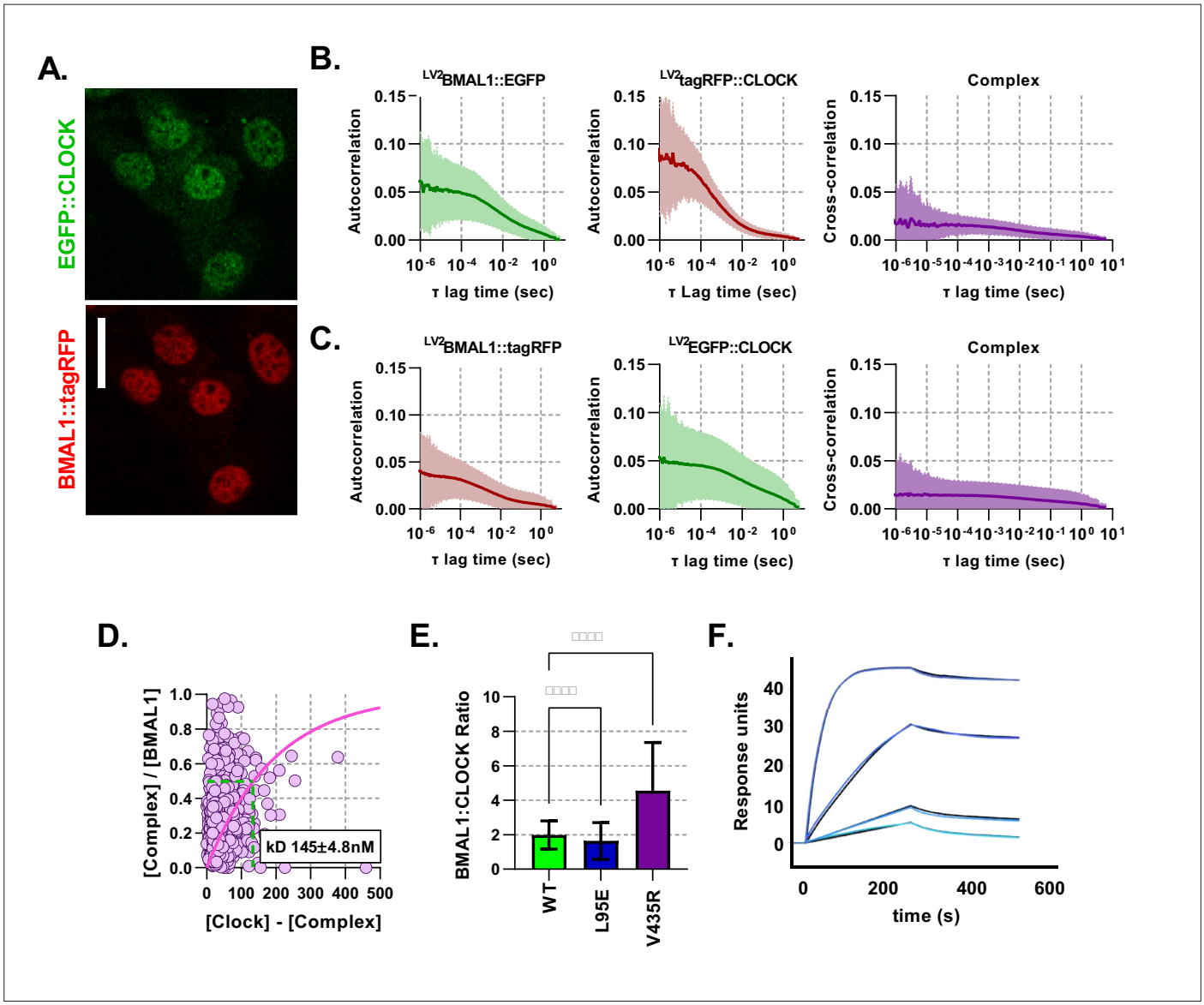

Figure 2—figure supplement 2. Fluorescent BMAL1 and CLOCK proteins behave similarly when colours are swapped. (**A**) Confocal images of NIH/3T3-LV2 EGFP::CLOCK-BMAL1::tagRFP cells. (**B**) Average auto- and cross- correlation curves shown as mean (line) and standard deviation (error envelope) for NIH/3T3 cells transduced to express BMAL1::EGFP and tagRFP::CLOCK. (**C**) Same as previous for colour swapped cells so that they express BMAL1::tagRFP and EGFP::CLOCK (n = 1158). (**D**) Dissociation plot to determine K_D for data from C. (**E**) Ratio calculations for number of nuclear molecules of BMAL1::EGFP/NLS::EGFP to tagRFP::CLOCK. (**F**) Surface plasmon resonance (SPR) analysis of heterodimer formation with immobilised biotinylated CLOCK PAS-AB in the presence of increasing concentrations of BMAL1 PAS-AB from 0.25 to 10 nm (light to dark blue). Data were fitted using a 1:1 binding model (global fit in black, association rate $k_{\mathrm{ON}} = 6.47 \times 10^5$, dissociation rate $k_{\mathrm{OFF}} = 8.98 \times 10^{-4}$, $K_D = 1.39 \times 10^{-9}$, $\chi^2 = 0.230$). Kruskal-Wallis test used to determine significance (values are denoted as p > 0.05 ns, p < 0.05 *, p < 0.01 **, p < 0.001 *** and p < 0.0001 ****).

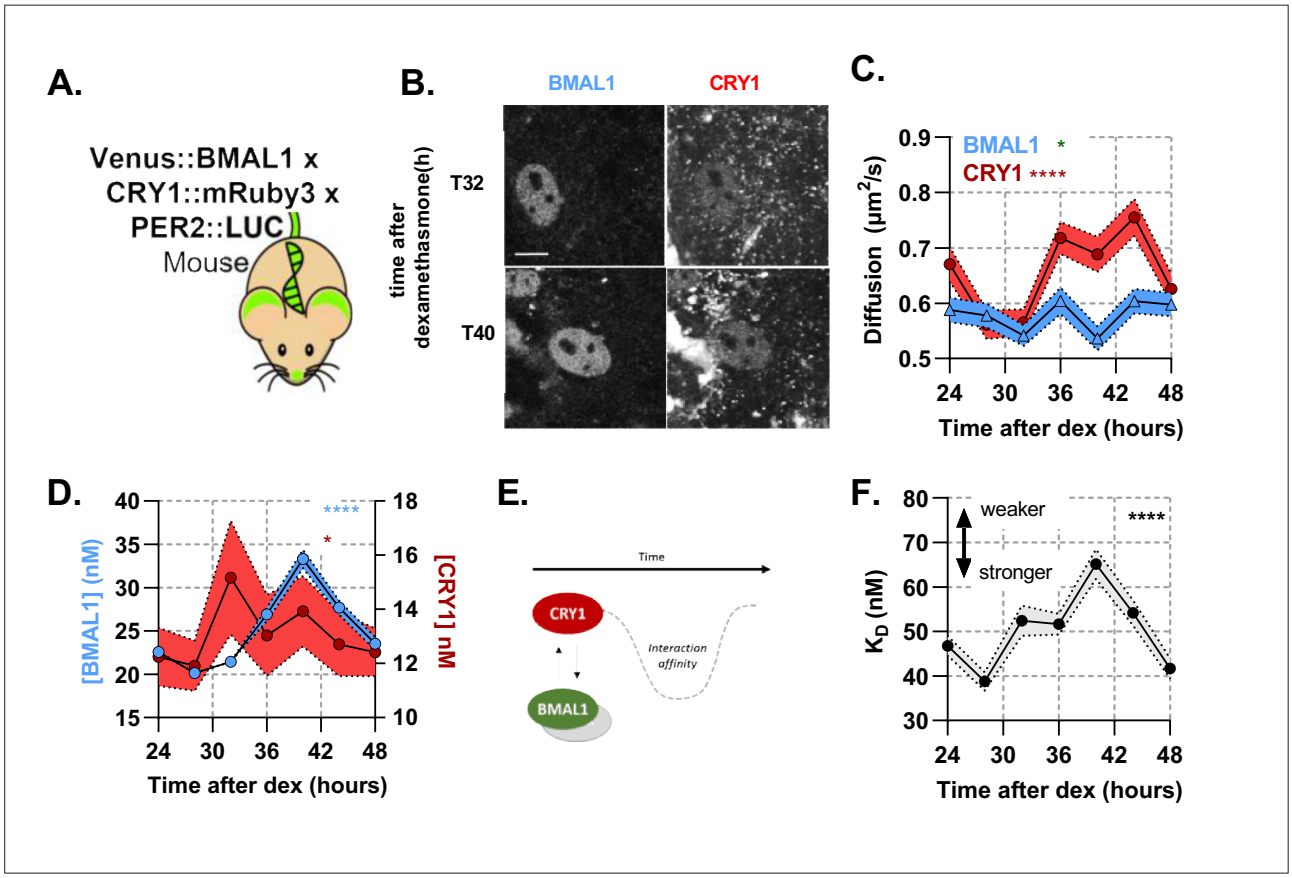

Figure 3. A rhythmic and strong interaction observed between slow-diffusing BMAL1 and CRY1 facilitates repression. (**A**) Schematic of triple-labelled mice from which primary lung fibroblasts were isolated (**B**) Confocal images of two cells shown for Venus::BMAL1 and CRY1::mRuby3 over time. FCS determined measurement for diffusion coefficient (**C**) and protein concentration (**D**) of Venus::BMAL1 and CRY1::mRuby3 (n = 136, 143, 173, 131, 158, 121, and 132; line shows the mean and error envelopes show the SEM). (**E–F**) Interaction strength between BMAL1 and CRY1 was also measured over time as illustrated by the schematic of affinity as well as plotted values of dissociation constant (error envelope shows the standard deviation). Kruskal-Wallis test used to determine significance (values are denoted as $p > 0.05$ ns, $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ *** and $p < 0.0001$ ****). Source data for panels B,C available as ***Figure 3—source data 1***, ***Figure 3—source data 2***, ***Figure 3—source data 3***, ***Figure 3—source data 4***.

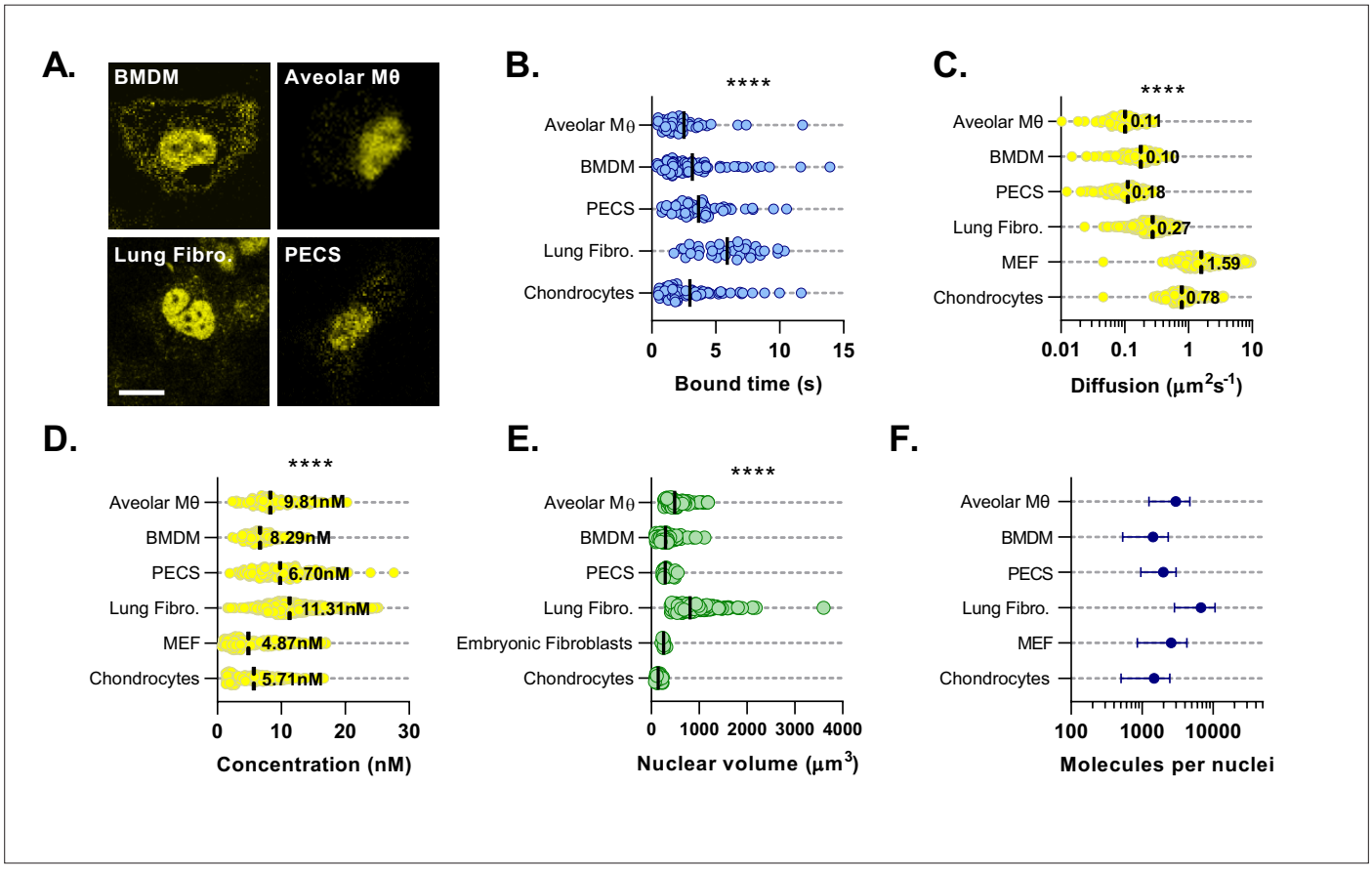

Figure 3—figure supplement 1. BMAL1 concentration and DNA binding parameters minimally vary across cell types. (**A**) Confocal microscopy images of primary cultures isolated from Venus::BMAL1 mice. (**B**) Characteristic bound time calculated from FRAP measurements of primary cultures from A. Chondrocyte data was previously measured by ***Yang et al., 2020*** and reanalysed for bound time (n = 73, 87, 61, 42, and 35 cells). (**C**) Cell cultures were measured using FCS and the auto-correlation data used to determine Venus::BMAL1 diffusion coefficient and (**D**) protein concentration (n = 107, 142, 156, 1597, 243, and 172 cells). (**E**) To determine total molecular abundance per nuclei, cultures were stained with Hoechst 33,342 and then imaged. Nuclear volumes were then determined (n = 84, 115, 27, 169, 9 and 30). (**F**) Total BMAL1 molecules calculated from average nuclear concentration and average nuclear volume. Kruskal-Wallis test was used to determine significance (values are denoted as $p > 0.05$ *, $p < 0.01$ **, $p < 0.001$ *** and $p < 0.0001$ ****). Panels B and D have been adapted from Figure 4E and 5A, B from ***Yang et al., 2020***.

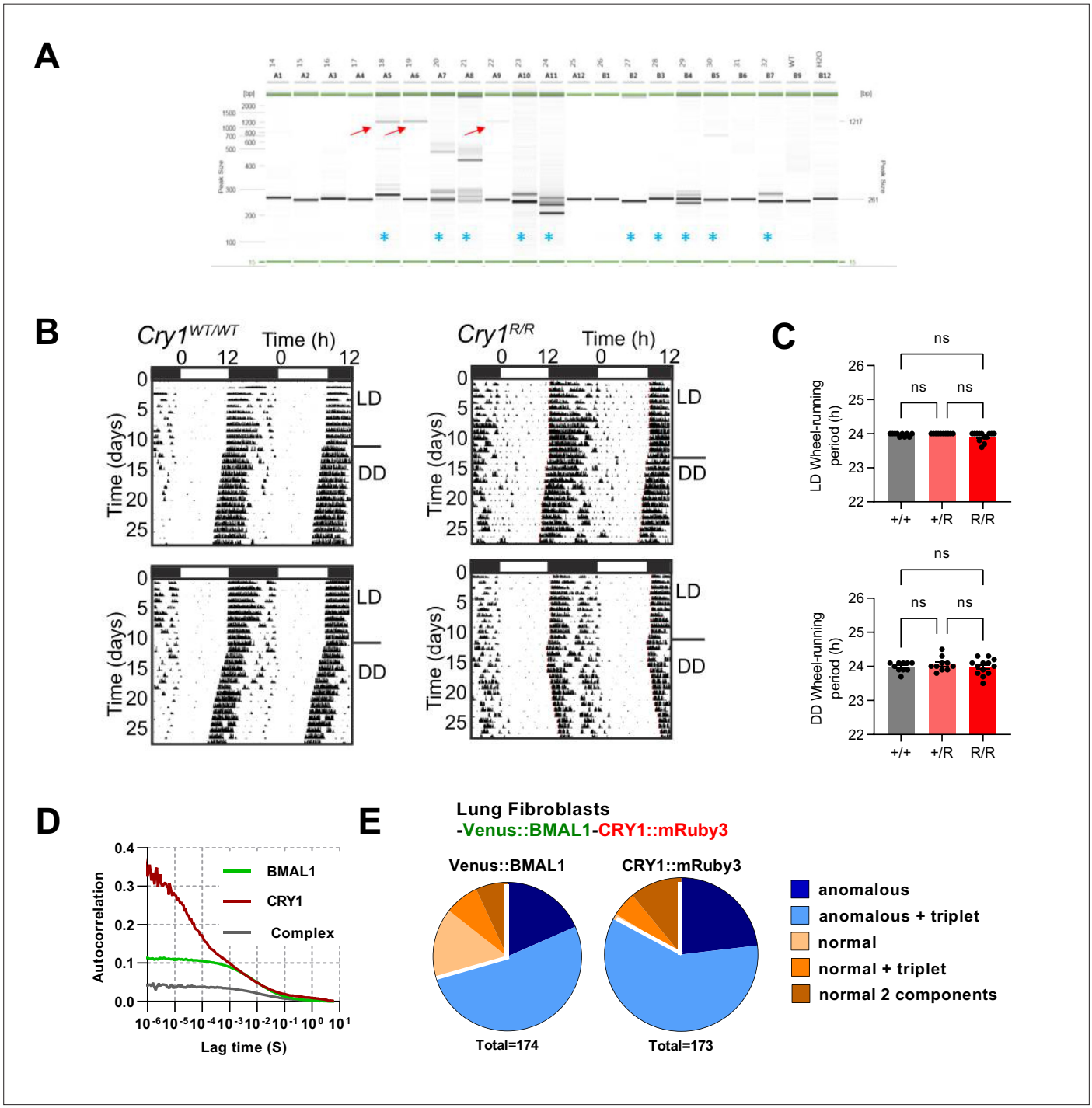

Figure 3—figure supplement 2. Generation of CRY1::mRuby3 mouse line. (**A**) Genotyping of pups by PCR. Images are PCR reactions run on Qiaxcel with red arrows indicating correct HDR product size, blue asterisks indicate mice in which InDels resulting from NHEJ are observed. (**B**) Actogram traces for wild-type and genetically modified CRY1::mRuby3 mice. (**C**) Mean ± SEM circadian periods for wheel-running in light-dark conditions (12 h/12 h) (+/+ = 10; +/*R* = 10; R/*R* = 13). Mean ± SEM circadian periods for wheel-running in constant dark (+/+ = 6; +/*R* = 10; R/*R* = 10). (**D**) Mean correlation curve for FCS measurements of BMAL1 x CRY1 x PER2::luc lung fibroblasts 24 hr after dexamethasone synchronisation (n = 144, BMAL1 and n = 135, CRY1). (**E**) FCS model selection results (pooling 24–48 hr post-dexamethasone measurements). One-way ANOVA test used to determine significance (values are denoted as $p > 0.05$ ns, $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ *** and $p < 0.0001$ ****).

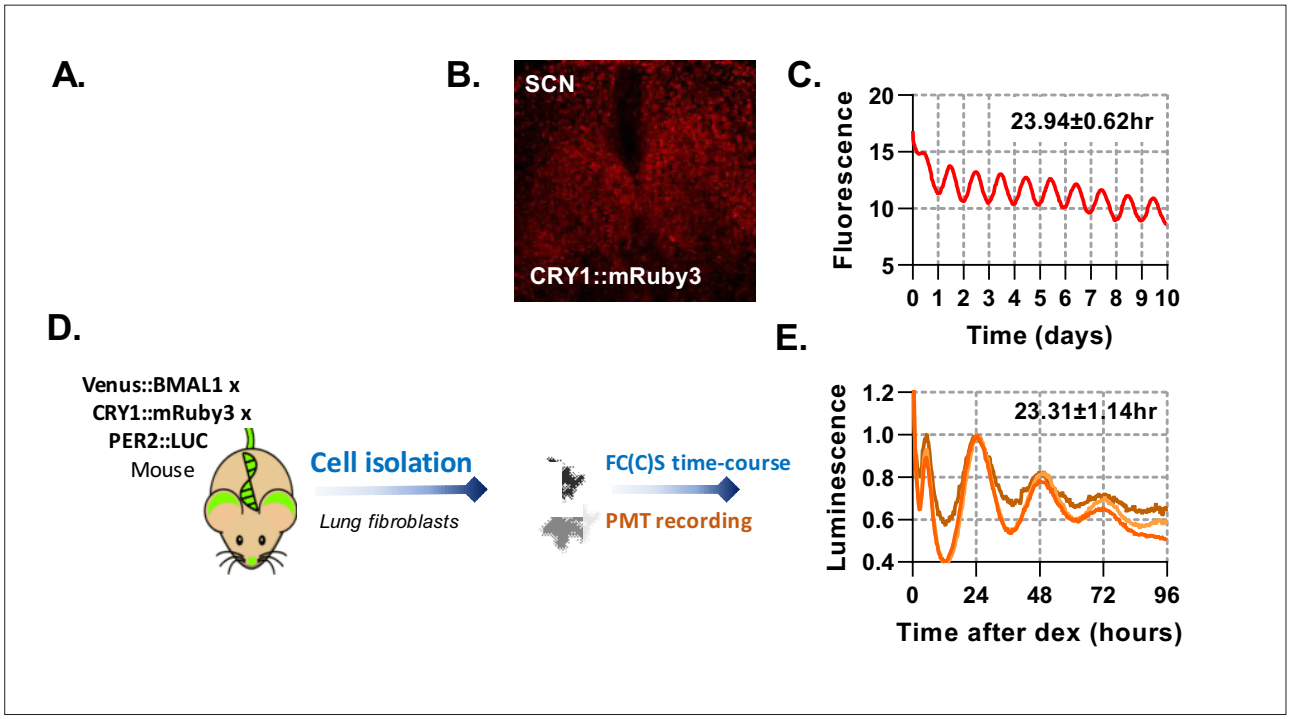

Figure 3—figure supplement 3. Triple endogenous labelled mice used to assay rhythms in SCN and peripheral lung fibroblasts. (**A**) Schematic representation of newly made transgenic mouse engineered to express CRY1::mRuby3, Venus::BMAL1, and PER2::LUC. (**B**) Confocal microscopy image of SCN organotypic slice expressing CRY1::mRuby3. (**C**) Quantification of mRuby3 fluorescence over time for the whole SCN, with mean and standard deviation of period. (**D**) Experimental set up to measure isolated primary lung fibroblasts from Venus::BMAL1 x CRY1::mRuby3 x PER2::LUC labelled mice synchronised with dexamethasone. Parallel cell cultures were analysed for luminescence and also by FCS over a time-course, measured every 4 hr. (**E**) Luminescence recordings of isolated primary lung fibroblasts from BMAL1 x CRY1 x PER2 labelled mice synchronised with dexamethasone. Data shown is for three independent replicates, with mean and standard deviation of period shown.

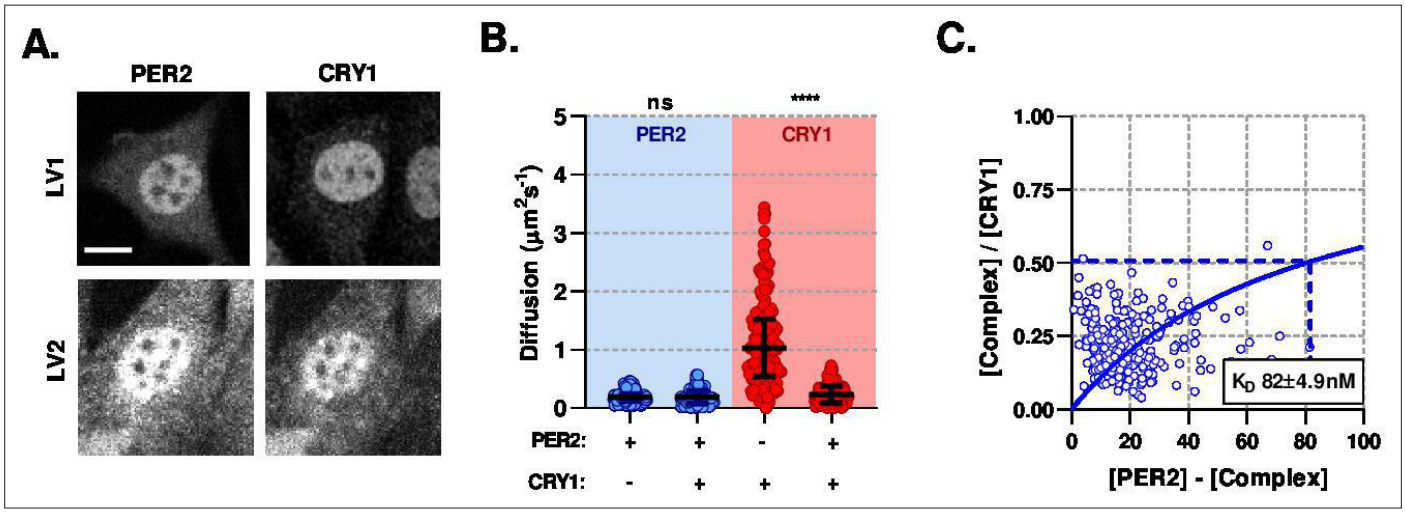

Figure 4. PER2 modulates CRY1 mobility via a high-affinity association. (**A**) Confocal images of transduced NIH/3T3 cells that either solo- or co- express PER2 and CRY1. (**B**) FCS data showing diffusion for PER2 and CRY1 in solo- and co-expressed conditions (n = 165, 174, 274, and 274 cells; diffusion rate means of 0.2, 0.2, 1.1, 0.2). (**C**) Dissociation plot from nuclear FCS measurements for EGFP::PER2 and CRY1::tagRFP (n = 274). Significance determined by Mann-Whitney test (values are denoted as $p > 0.05$ ns, $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ *** and $p < 0.0001$ ****). Source data for panels B available as ***Figure 4—source data 1***.

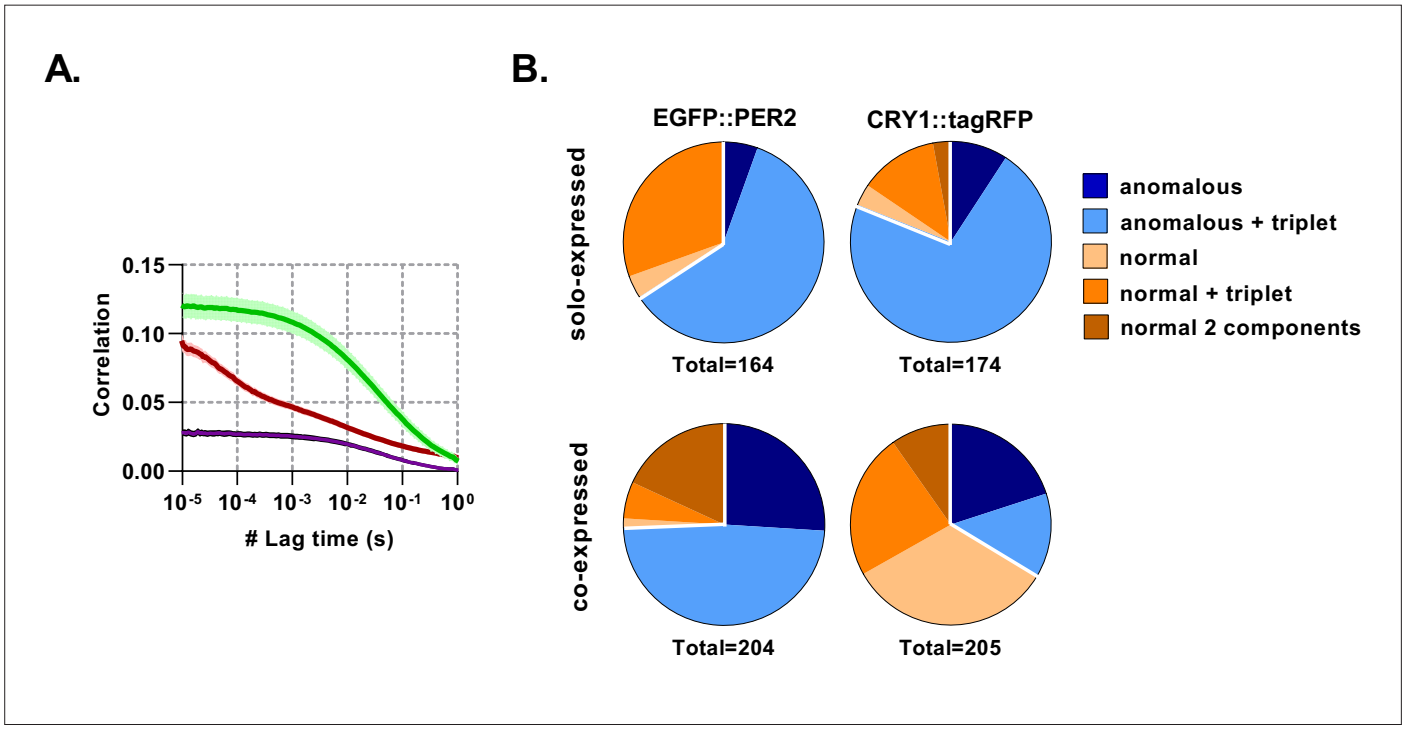

Figure 4—figure supplement 1. CRY1 mobility is affected by co-expression with PER2. (**A**) Average auto- and cross- correlation curves shown as mean (line) and standard deviation (error envelope) for NIH/3T3 cells transduced to express EGFP::PER2 and CRY1::tagRFP. Measurements were made in the nuclei. (**B**) FCS model selection results for cells that either solo-express CRY1/PER2 or co-express both.

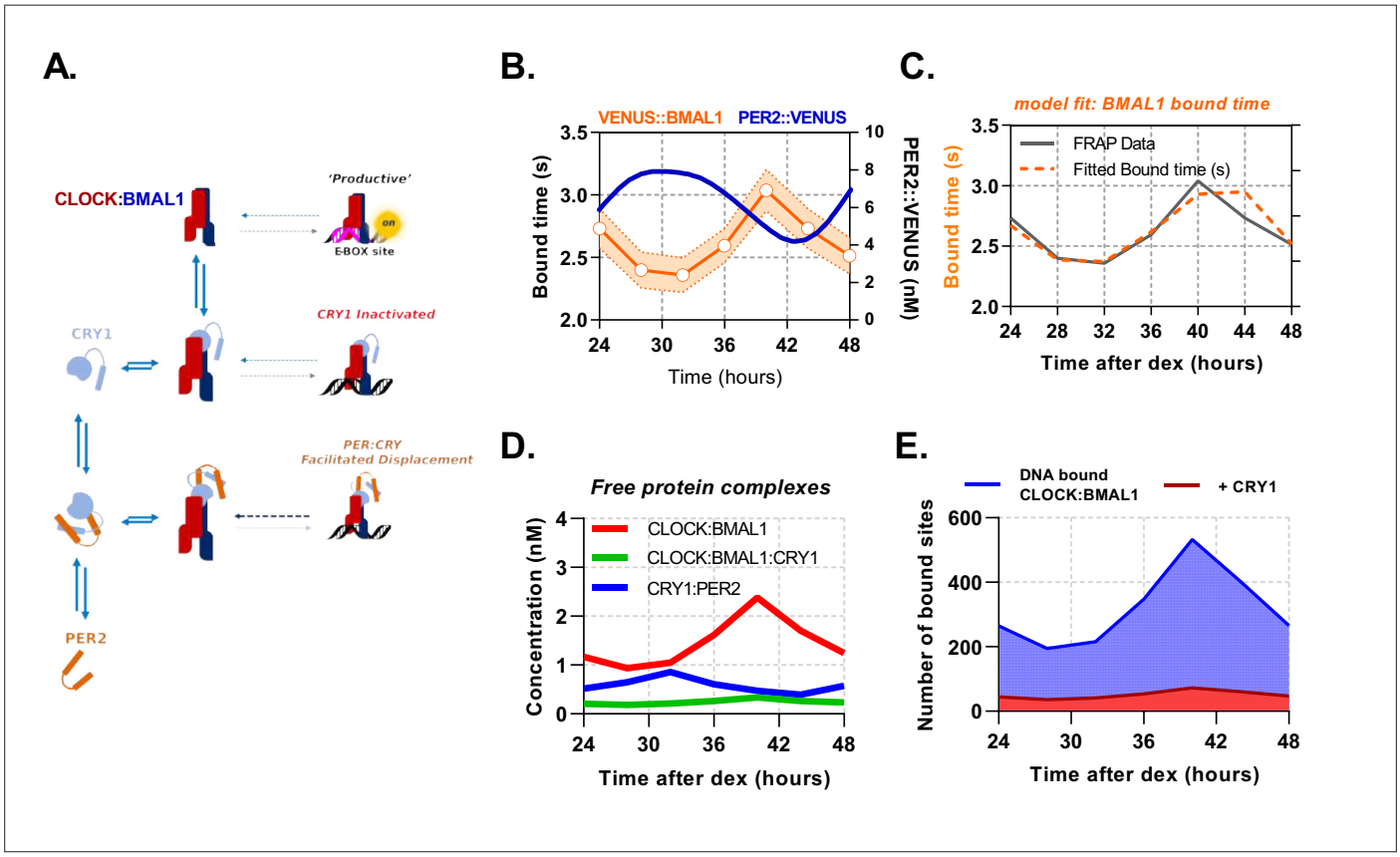

Figure 5. PER2 acts via CRY1 to mediate rhythmic displacement of CLOCK:BMAL1 from DNA. (**A**) Schematic representation of model topology used for the deterministic model of CLOCK:BMAL1 DNA binding. (**B**) Primary lung fibroblasts from BMAL1 x CRY1 x PER2 mice were synchronised with dexamethasone. Plot shows PER2 concentration as measured via FCS by ***Smyllie et al., 2016*** as well as mean BMAL1 binding time (showing SEM error envelope). Binding time was measured by confocal FRAP measurements performed on the Venus::BMAL1 fluorescence. Orange line shows the inverse of kOFF (s^{-1}), determined from fitting the recovery data with a single component model (n = 48, 70, 82, 63, 82, 64, and 65 cells). (**C**) ODE model was fit to FRAP binding data from E. and using a measured input for PER2 nuclear concentration previously determined in ***Smyllie et al., 2016***. Model output showing (**D**) inferred nuclear concentrations for molecular complexes (**E**) and CLOCK:BMAL1 without and with CRY1 bound to target sites (see supplementary materials for parameters). Panel B has been adapted from Figure 3C from ***Smyllie et al., 2016***.

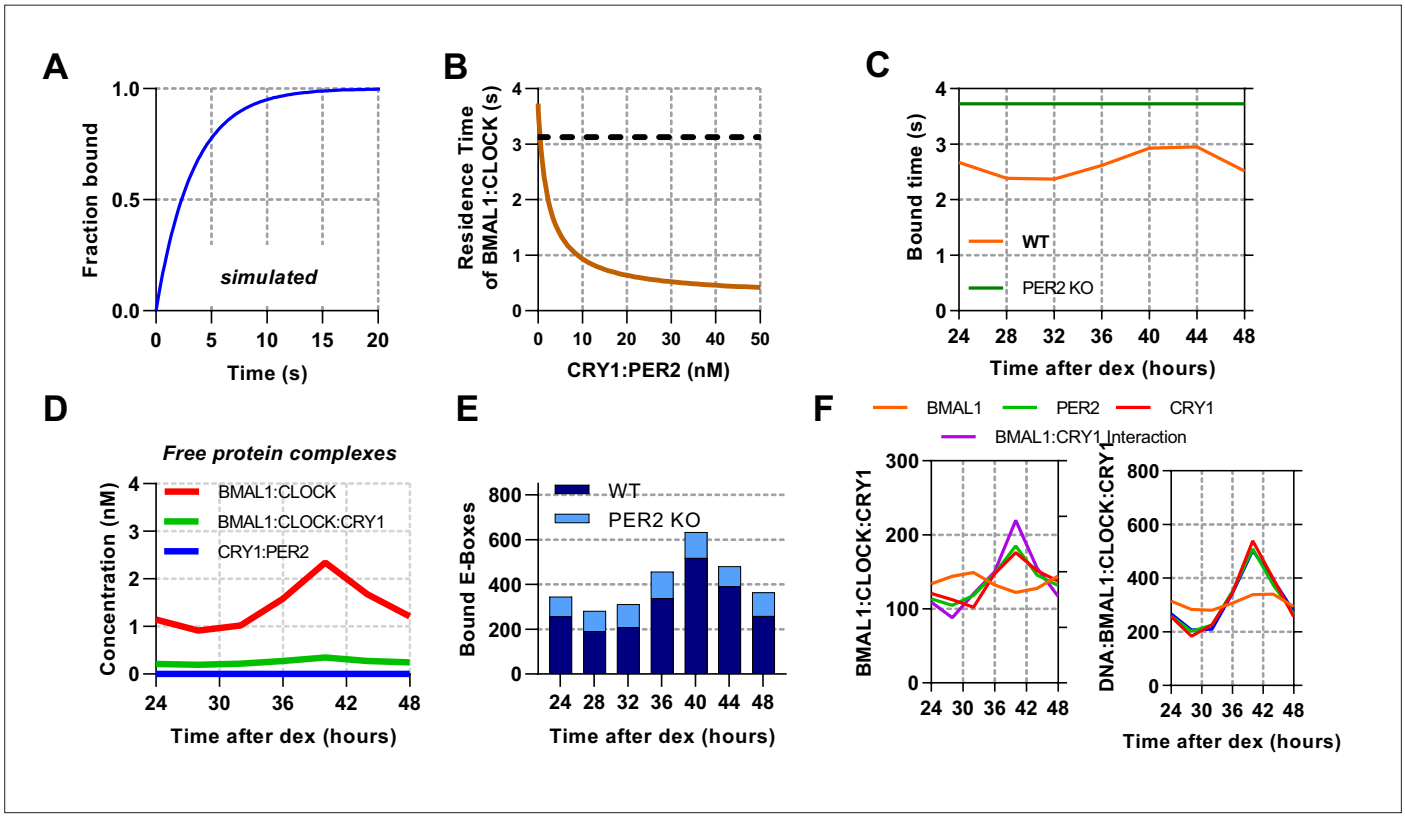

Figure 5—figure supplement 1. ODE model of CLOCK:BMAL1 DNA binding using measured inputs and modelled perturbations. (**A**) Normalised model simulated FRAP using ODE model by counting recovery of site bound BMAL1 species after removal. (**B**) Model determined residence time of DNA bound CLOCK:BMAL1 across different concentrations of CRY1:PER2 using parameters from T40. (**C**) Modelled residence time of DNA bound CLOCK:BMAL1 across circadian time calculated for WT and without PER2. Without PER2 model data over a circadian cycle showing (**D**) non DNA-bound complexes and (**E**) DNA bound CLOCK:BMAL1. (**F**) Plots showing the CLOCK:BMAL1:CRY1 complex across several simulated conditions, including removal of rhythmicity from BMAL1, CRY1, or PER2 protein levels or the interaction between BMAL1:CRY1 (fixing them to their mean concentration).

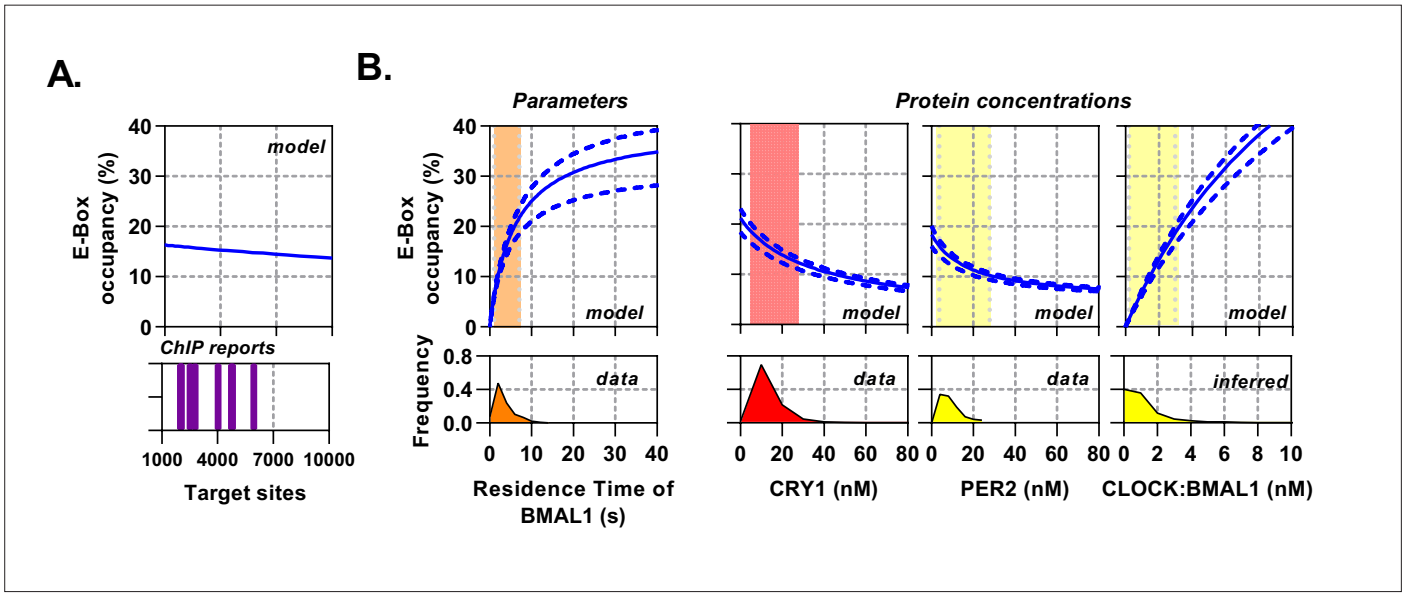

Figure 6. Circadian proteins operate within an optimal range to modulate E-Box binding. Sensitivity analysis of the deterministic binding model showing relationship of measured parameters (bottom) against model for occupancy of active BMAL1:CLOCK on target sites (top). (**A**) Changing number of target sites with data matched to BMAL1 ChIP data sets. (**B**) From left to right, the effect of changing residence time of CLOCK:BMAL1, or protein concentrations. Histograms show measured concentrations for corresponding proteins across all conditions/cells. The 10th to 90th percentile is highlighted. Source data available as ***Figure 6—source data 1***.

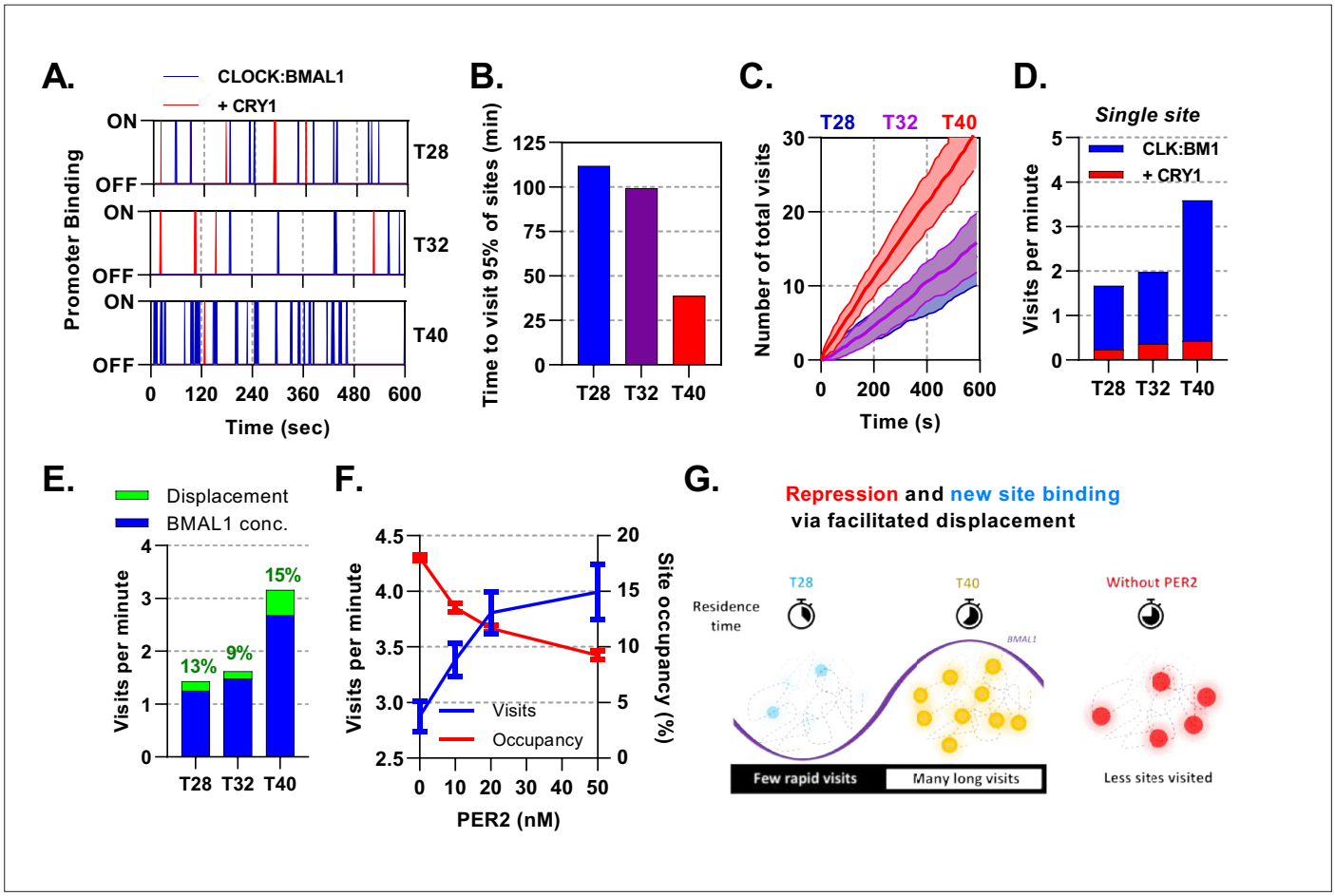

Figure 7. Mathematical modelling demonstrates dual function of PER:CRY mediated repression. Stochastic binding model outputs using parameters corresponding to T28, T32 or T40 post dexamethasone BMAL1 x CRY1 data sets. (**A**) Shows a promoter corresponding to the average binding rate of CLOCK:BMAL1, (**B**) the time to visit 95% of target sites once and (**C**) number of visits to a single promoter over time. Shaded error envelope shows standard deviation. (**D**) Average number of visits per minute to a target site showing active and CRY1 repressed CLOCK:BMAL1 visits. (**E**). Comparison of the contribution of BMAL1 concentration (blue) and PER2 facilitated displacement (green) on the visits per minute to a target site. Percentage contribution indicated. (**F**) Relationship of PER2 protein concentration to site visitations per minute and occupancy by CLOCK:BMAL1 using parameters for T40 explored over different concentrations of PER2. Error bars represent standard deviation. (**G**) The action of CRY:PER leads to short-lived transient binding of CLOCK:BMAL1 to DNA, working as both a repressive action whilst also facilitating binding to new target sites.

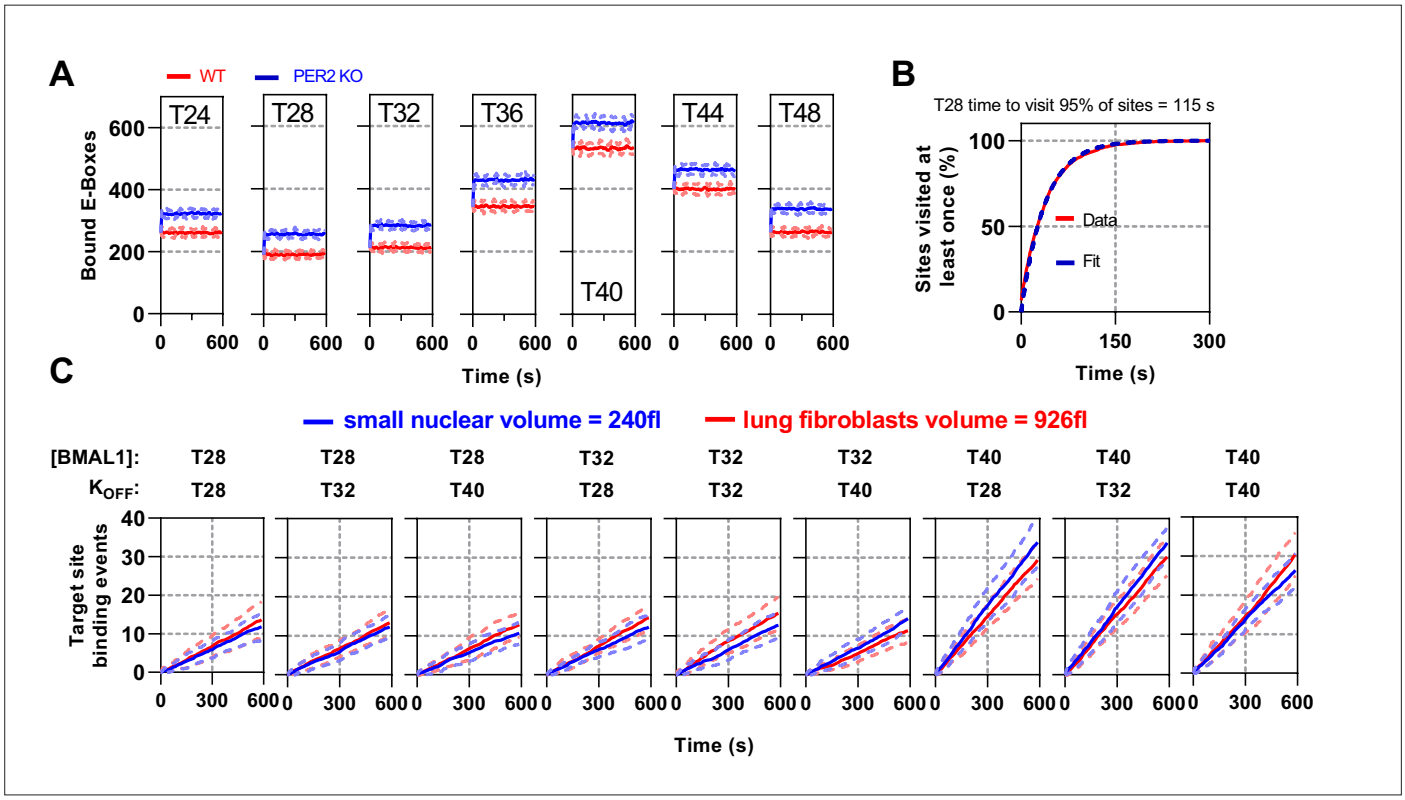

Figure 7—figure supplement 1. Stochastic binding model using experimentally measured parameters (**A**) Stochastic model showing the average binding (with SD) of CLOCK:BMAL1 bound target sites using input measurements from all time points for both WT and without PER2 simulations. (**B**) The time to visit every E-Box site once for T28 showing fit. (**C**) Model simulation plots showing CLOCK:BMAL1 visits to a single promoter over time. Red and blue lines show simulations using different nuclear volumes (red and blue lines, corresponding to primary lung fibroblasts and mouse embryonic fibroblasts). Input values for BMAL1 concentration and binding OFF rate correspond to those measured at the time point referred above each panel.

Chapter 3

Protein-protein interactions of the circadian clock

3.1 Journal paper: Anatomy of circadian clock protein complexes in live cells

Title	Anatomy of circadian clock protein complexes in live cells
Journal	Manuscript in preparation
Year	N/A
Authors	James S Bagnall*, **Alex A Koch***, Nicola Begley, Nicola J Smyllie, Antony D Adamson, David G Spiller, Carrie L Partch, Korbinian Strimmer, Thomas A House, Michael H Hastings, Andrew SI Loudon
doi	N/A

* These authors contributed equally

3.1.1 My contributions

Alongside James Bagnall and Andrew Loudon, I was part of the development and write up of this paper. I assisted with the experimental work carried out by James Bagnall and Nicola Begley. Finally, I developed the new robust approach to FCCS analysis as well as the mathematical modelling of facilitated interactions.

Article

Anatomy of circadian clock protein complexes in live cells

James S. Bagnall[1, †], Alex A. Koch[1, †], Nicola Begley[1], Nicola Smyllie[2], Antony Adamson[1], David G. Spiller[1], Carrie L. Partch[3], Korbinian Strimmer[4], Thomas A. House[4], Michael H. Hastings[2], and Andrew S. I. Loudon[1, *]

[1]Faculty of Biology, Medicine and Health, University of Manchester, UK, [2]MRC Laboratory of Molecular Biology, Cambridge, UK, [3]Department of Chemistry and Biochemistry, University of California, Santa Cruz, USA, [4]Department of Mathematics, University of Manchester, UK, [*]Correspondence: andrew.loudon@manchester.ac.uk, [†]These authors contributed equally to this work

Abstract

Circadian rhythms (~24 hour) play a critical role in regulating a variety of physiological functions across time, and are driven by a cell-autonomous clock that relies on multiple protein-protein interactions. Yet, how these interactions come together in time and space within living cells is largely unexplored. Here, we quantified the dynamic interactions between 15 pairs of core circadian clock proteins in live cells, using Fluorescence Cross-Correlation Spectroscopy (FCCS). Remarkably, we found that the protein pairs PER2-CRY1 and CRY1-BMAL1, were the strongest interactors with very similar dissociation constants (K_D). Moreover, the strengths of these interactions were well-conserved across different cell types and protein expression levels. We developed new methodologies to establish which interactions are direct or facilitated by a third partner, finding that most proteins in the circadian network directly bind with one another. In a limited number of cases, PER2 and BMAL1 were found to play a central role in facilitating complex formation, acting as scaffolds for multi-protein complexes involved in the repressive phase of the circadian cycle. We validated our findings by measuring three-way interactions by FCCS. From this, we show that the strong PER2-CRY1 and PER2-CRY2 interactions can stabilise CRY proteins and are sufficient to maintain post-translational rhythms of CRY2, independently of transcription. This study provides new insight into how multiple protein components interact in time and space within the living cell and offers a general quantitative approach applicable to other biological systems.

Introduction

The circadian clock is a fundamental adaptive feature for virtually all living organisms. Circadian clocks are cell-autonomous, and in mammals, organise multiple physiological functions, as diverse as immunity (Scheiermann et al., 2018), metabolism (Buhr and Takahashi, 2013), and behaviour (Konopka and Benzer, 1971).

The mammalian circadian system is coordinated in a hierarchical manner. Retinal photic input entrains a specialised hypothalamic structure (suprachiasmatic nucleus, SCN), outflow from which impacts on multiple peripheral autonomous oscillators throughout the body through a diverse range of cues (Meyer-Bernstein et al., 1999; Buijs et al., 2003). At the cellular level, a molecular network of interconnected feedback loops drives rhythmic oscillations in a small number of canonical clock genes, in a process defined as a transcription-translation feedback loop (TTFL). At the centre of the TTFL is the heterodimeric transcription factor comprised of the proteins Circadian Locomotor Output Cycles protein Kaput (CLOCK) and Brain and Muscle ARNT-like 1 (BMAL1). Genetic disruption of either of these proteins leads to abnormal circadian function and arrhythmia (Vitaterna et al., 1994; Haque et al., 2019). CLOCK:BMAL1 regulate gene expression of core rhythmic negative feedback protein regulators of the circadian clock, PERIODs (PER1-3) and CRYPTOCHROMEs (CRY1-2), with ancillary loops encoded by Rev-Erbα/β and RAR-related orphan receptors (ROR$\alpha-\gamma$). Once translated into their respective proteins, PERs and CRYs act to inhibit their own transcription by forming multimeric protein complexes encompassing many components, including PER1-3, CRY1/2, Casein Kinases 1δ/ϵ (CK1δ/ϵ), CLOCK, and BMAL1 (Aryal et al., 2017; Cao et al., 2023). Critically, nuclear translocation requires dimerization of PER and CRY proteins and in the absence of CRYs, PERs are unable to repress CLOCK:BMAL1 (Smyllie et al., 2022). Once in the nucleus, repression by CRYs and PERs proceeds in two modes, "blocking" and "displacement" type repressions (Partch, 2020). In the initial repressive phase, PER:CRY complexes bind to, and remove, CLOCK:BMAL1 from DNA by initiating phosphorylation of CLOCK:BMAL1 by CK1δ/ϵ ("displacement" type) (Cao et al., 2021, 2023; Koch et al., 2022). In the second stage, CRY1 competitively binds to, and inhibits, the c-terminal transactivation domain of BMAL1, thereby blocking recruitment of co-factors such as CBP and p300 required for initiation of transcription (Xu et al., 2015). Degradation of CRYs and PERs leads to alleviation of repression, and initiation of active CLOCK:BMAL1

mediated transcription (Koike et al., 2012).

Protein-protein interactions described above are known to be crucial in the functioning of the circadian clock. A number of studies have characterised protein interactions using in-vitro methods, including immunoblotting and autoradiograms (Cao et al., 2021), isothermal titration calorimetry, fluorescence polarization (Xu et al., 2015) and yeast two-hybrid systems (Wallach et al., 2013). A single particle electron microscopy study by Aryal et al. (2017) has suggested these complexes incorporate multiple proteins simultaneously in large megadalton macromolecular complexes. However, a technical difficulty with all of these approaches is that protein-protein interactions are studied in vitro or within fixed tissues and they lack details of the behaviour of proteins in living cells, sub-cellular localisations, normal post-translational modifications, or insight into protein temporal dynamics over the course of the circadian cycle. To address this, we previously used the advanced microscopy technique of Fluorescence Cross-Correlation Spectroscopy (FCCS). This method permits the study of protein complex formation in its biological context non-destructively over time (Bacia et al., 2006; Sadaie et al., 2014) and from this we have measured two spectrally distinct fluorescently labelled proteins (BMAL1 and CRY1) in order to define protein-protein complex formation in living cells (Koch et al., 2022).

Here, we comprehensively extend these approaches to define interactions of all core proteins of the circadian TTFL clockwork mechanism, encompassing 15 protein pairs and 6 monomers. Many of these interactions were found to be strong, rhythmic, and conserved across different cell types as well as expression levels. In order to extend this to multiple protein complexes, we developed mathematical models and measurement of three-way FCCS. Strikingly, this demonstrated that the circadian clock primarily operates via a temporally organised series of pair-wise protein-protein interactions, with only a small number of complexes facilitated by an additional third partner. By measuring three proteins simultaneously, we further characterised the impact of complex formation on each constituent protein, and from this showed that stabilisation by PER protein of CRY is sufficient to generate rhythmicity, even in the absence of rhythmic transcription. Here, our live-cell methodologies provide new insights into how multiple proteins form complexes within in time and space as well as presenting a general quantitative approach to investigate other biological systems.

Results

Live-cell measurement of interaction between PER2 and CRY1 in primary fibroblasts

We employed FCCS, to assess the molecular properties of multiple fluorescent fusion proteins in living cells (Bacia et al., 2006; Komatsubara et al., 2019). Issues can arise in FCCS, as the two differently coloured fluorescent proteins may be subject to different illumination volumes (Figure supplement 1), as well as differences in photochemical properties, such as triplet states, blinking, bleaching sensitivities, and maturation rates (Koch et al., 2022; Hess and Webb, 2002; Schönle et al., 2014; Lambert, 2019). To account for this, we first calibrated FCCS by measuring cells which expressed enhanced green fluorescent protein (EGFP) fused with the red fluorescent protein tagRFP (Figure supplement 1A), so that each expressed protein has both fluorescent moities. Thus, any discrepancies in observed concentrations and diffusion rates should be due to differences in their respective excitation confocal volumes or fluorescent photo-physical properties (Figure supplement 1B). We developed a correction, dependent on excitation wavelength (see Equation 1) Materials and Methods) for differences in confocal volumes, and thereby concentration (Figure supplement 1C vs D) and diffusion (Figure supplement 1E). These corrected estimates demonstrated a consistent 1:1 relationship, confirming the accuracy and suitability of this approach for subsequent use.

PER2 and CRY1 are well-defined circadian proteins involved in rhythmic repression, and yet their interaction across circadian time in living cells was untested. To address this, we utilised the previously established CRISPR-modified double knock-in PER2::Venus x CRY1::mRuby3 mouse (Koch et al., 2022; Smyllie et al., 2022). Using skin fibroblasts in culture, which were then circadian-synchronised using dexamethasone (DEX), we undertook a time-course study 24-48h post-DEX synchronisation, with FCCS measurements every four hours (Figure 1A, B). Within the nucleus proteins diffuse slower than expected by classical Brownian motion due to the highly crowded and heterogenous environment. The movement of these proteins is primarily constrained by interactions with other molecules, including DNA. Thus, FCCS auto-correlation data were analysed using a previously described anomalous diffusion model that captures the sub-diffusive behaviour of nuclear proteins (Koch et al., 2022).

Owing to cell-to-cell variation and uncertainty inherent to FCCS analysis, robust quantification requires many repeat measures, as shown for example in Figure 1C. Here, interaction strength was quantified as a dissociation constant K_D, an inverse measure of complex formation that represents the concentration at which 50% of the partner proteins will be in complex. Hence, a lower K_D represents a stronger interaction. K_D was determined from a linear fit obtained by quantile regression, relating the concentration of protein in complex, [PER2:CRY1], to the concentrations of unbound constituent proteins, $[PER2]_{Free}$ and $[CRY1]_{Free}$ (Figure 1D). We denote complexes between proteins as PER2:CRY1 showing that PER2 protein is bound with CRY1. Analysis of the time-course revealed PER2 protein to be more abundant than CRY1, by approximately 2-fold at peak. PER2 protein

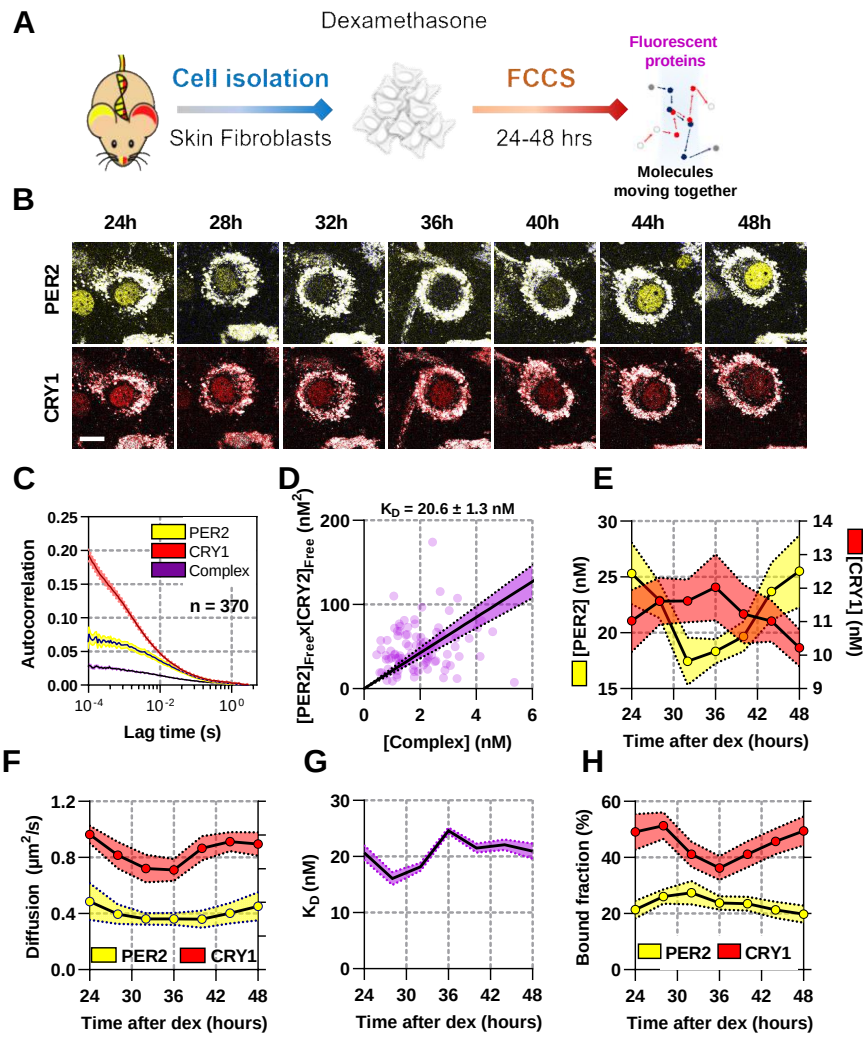

Figure 1: FCCS Measurements of endogenous PER2 and CRY1 protein molecules in primary skin fibroblasts isolated from knock-in mice. (**A**) Skin fibroblasts were isolated from knock-in mice expressing fluorescent PER2::Venus and CRY1::mRuby3. Cells were then stimulated with a 1-hour dexamethasone shock prior to imaging and FCCS measurements. (**B**) Confocal microscopy images of a cell across 24 hours expressing nuclear PER2 and CRY1. Images show a merge of signal (yellow or red) and an autofluorescence channel (white). FCCS measurements were performed in the nucleus across >100 cells per time point every four hours, with biological replicates coming from cells isolated from four separate mice. (**C**) Example average FCCS correlation data for a single time-point 24-hours post-DEX made in cells labelled with PER2::Venus (yellow) and CRY1::mRuby3 (red). Overlaid black lines show the fits to these data used to calculate concentrations and diffusion coefficients. When proteins interact and form a complex, they move together such that their signals are well correlated yielding a non-zero cross-correlation curve (purple). (**D**) Strength of complex formation is measured by the ratio of the concentration of free proteins to complex, known as the disassociation constant (K_D). K_D is calculated as the slope of a linear fit between the product of free protein concentrations and complex, here shown for the 24-hour post-DEX time-point. Plots show measurements for median fluorescent (**E**) protein concentration with 95% confidence envelope, (**F**) diffusion and (**G**) disassociation constant with 95% confidence envelope displayed. Concentration (**E**) and diffusion rates (**F**) for both PER2 and CRY1 oscillated significantly, determined from ANOVA p-values of < 0.001. (**H**) Percentage of total PER2::Venus and CRY1::mRuby3 proteins bound in complex.

abundance cycled with a 2-fold amplitude of over 24h. In contrast CRY1 exhibited a lower amplitude cycle (Figure 1E). Measures of protein mobility (diffusion rate) showed PER2 to move on average, at half the rate of CRY1 overall (Figure 1F), and slower than our previous reports for BMAL1 (Koch et al., 2022). When mutually bound in complex, PER2 and CRY1 will diffuse at the same rate, however our FCS measurements of diffusion rates are averages for all PER2 and CRY1 proteins in nucleus and thus include faster diffusing free proteins. We next assessed live-cell interaction

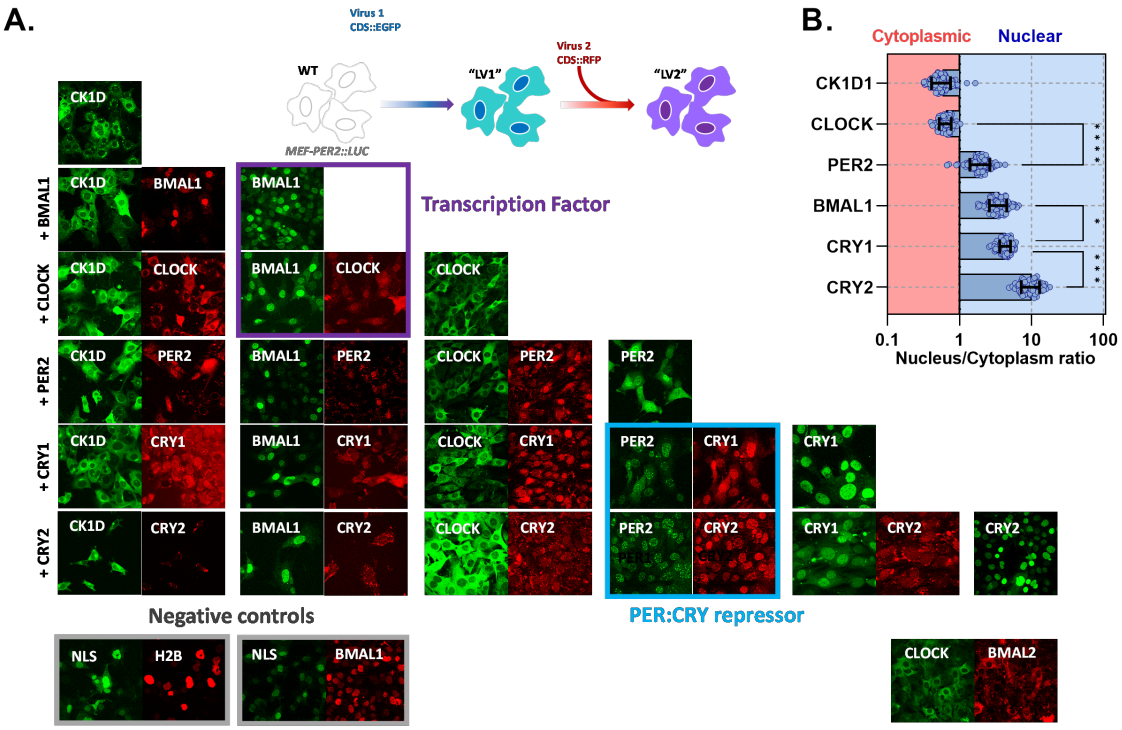

Figure 2: Expanding analyses to pairwise cell models (LV2) to study localisation and live-cell interactions of circadian proteins. (**A**) Mouse embryonic fibroblasts (MEFs) are either singularly or sequentially transduced to express fluorescent fusions of key circadian proteins with either EGFP or tagRFP. A nuclear localised EGFP (NLS) and Histone H2B::tagRFP were used as negative controls. Single transduced cells are denoted as LV1 and co-expressed as LV2. Representative confocal microscopy images in asynchronous timing conditions of MEFs either solo-expressing EGFP (LV1) or co-expressing EGFP and tagRFP fluorescent proteins (LV2). (**B**) Quantification of nuclear to cytoplasmic ratio (N:C) in asynchronous timing conditions from cells expressing a single fluorescent protein ordered from most cytoplasmic to most nuclear. Black bars indicate standard deviation. Majority of solo-expressed proteins were predominately nuclear, with key exceptions for the cytoplasmic localisations of CK1δ and CLOCK. Notably, under co-expression with BMAL1, CLOCK is localised into the nucleus. Statistical analysis was carried out using Kruskal-Wallis (values are denoted as $p < 0.05$ *, $p < 0.001$ *** and $p < 0.0001$ ****).

data over time revealing a very tight association between PER2 and CRY1, with an average K_D of 20 nM, with a significant but low amplitude cycle of 20% (Figure 1G). The proportion of proteins bound over 24h was approximately 50% vs 20% for CRY1 and PER2 respectively, (Figure 1H) . For comparison, we measured non-synchronised (i.e. non-DEX treated cells) and found similar values for both diffusion and interaction compared with circadian synchronised cells (Figure supplement 2). Additionally, the steady state concentrations of PER2 and CRY1 were within the same range as those observed following circadian synchronisation (Figure supplement 2F). Thus, PER2 and CRY1 exhibit a circadian anti-phasic relationship for abundance and the proportion of protein bound in complex.

Over and out of context-expression of paired fluorescent proteins to study live-cell protein-protein interactions of the circadian clock

CRISPR engineering of multiple modified lines of mice to knock-in fluorescent labels within the endogenous locus of many pairs of proteins is expensive and time consuming. To overcome this hurdle, we developed a method to assay multiple paired interactions of core components of the circadian clock using a lentiviral expression system. All these fluorescent fusion proteins were expressed from the non-circadian Ubiquitin C (UbC) promoter. Using mouse embryonic cells we established an initial series of 7 lines, transduced to express PER2, CRY1/2, BMAL1/2, CLOCK and CK1δ labelled with EGFP, denoted as LV1 (labelled proteins use the nomenclature :: to denote fusion i.e. BMAL1::EGFP; Figure 2A). Appropriate sub-cellular localisation of each fluorescent fusion was assessed by measuring the nuclear/cytoplasmic (N:C) ratio of fluorescence. This showed CRY1 and CRY2 to be strongly nuclear, (with CRY2 most nuclear), confirming our

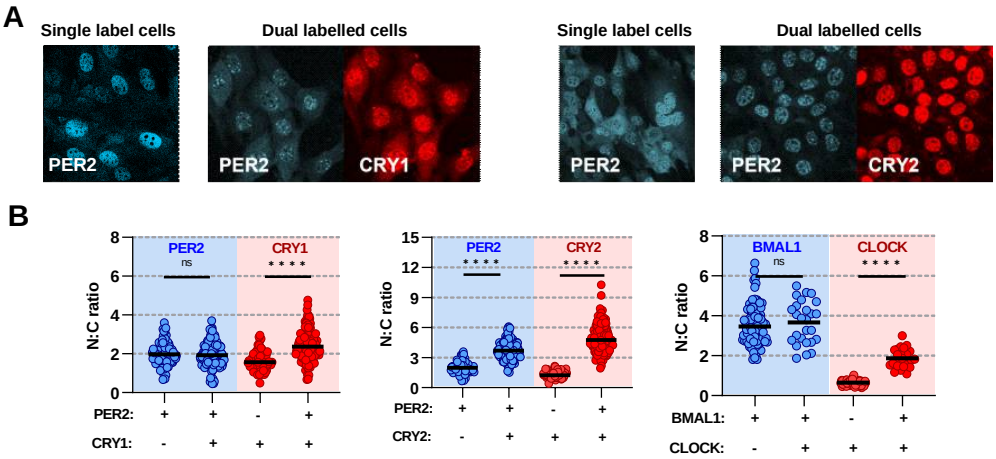

Figure 3: CRY1/2 and CLOCK localisation is dependent on expression of interaction partners. (**A**) Confocal microscopy images of MEFs transduced to express PER2 and CRY paralogs alone (single panel images) or together (dual panel images). (**B**) Quantification of nuclear to cytoplasmic (N:C) ratio calculated by the intensity of nuclear fluorescence divided by cytoplasmic fluorescence across single and dual labelled conditions. Black bar indicates median value. Blue boxes denote N:C ratio for EGFP labelled proteins and red boxes correspond to proteins fused to tagRFP. Statistical analysis was carried out using one-way ANOVA (values are denoted as $p < 0.0001$ ****).

prior observations in SCN organotypic slices (Figure supplement 3A) (Smyllie et al., 2022). BMAL1 and PER2 were also primarily nuclear in expression, consistent with our previous findings across multiple cell types (Smyllie et al., 2022; Yang et al., 2020; Koch et al., 2022) but with a lower N:C ratio than for CRY proteins. In contrast, CLOCK, BMAL2, and the pace setting kinase CK1δ were mostly cytoplasmic when over-expressed in isolation (Figure 2B).

Next, we introduced additional proteins fused with the red fluorescent protein tagRFP to 6 of the above cell lines by a second transduction, establishing a further 15 lines covering all pairwise combinations, and denoted as LV2 (Figure 2A, Figure supplement 3B). Two additional non-interacting negative control cell lines were established for nuclear localised EGFP (NLS) with either Histone H2B (H2B) or BMAL1 fused with tagRFP (Figure 2A). As FCCS measures interactions in a small sub-cellular volume, it was important to determine the localisation of each pair of proteins, and where they may interact prior to measurement. By these means, we extended our nuclear/cytoplasmic fluorescence assay to every fluorescent pair. This showed that some proteins change localisation upon co-expression (Figure supplement 4). For example, nuclear localisation of both CRY1 and CRY2 was strongly dependent on PER2. Since differences in autofluorescence backgrounds across EGFP and tagRFP recordings makes comparison of N:C ratio across two colours difficult, we produced an additional 2 single labelled LV1 cell lines expressing red labelled CRY1 and CRY2 without a green partner to compare with the previous cells expressing EGFP labelled PER1/2 and red CRY1/2 (Figure 3A). This clearly showed a significant increase in nuclear CRY1 and CRY2 in the presence of PER2, with CRY2 demonstrating a striking 5-fold increase in nuclear localisation (Figure 3B). Additionally, we show that CLOCK is a cytoplasmic protein, unless expressed with BMAL1 (Figure 3B), a recapitulation of our previous findings using NIH/3T3 fibroblast cells (Koch et al., 2022).

We next measured the affinity of interaction (K_D) between protein pairs as assessed by FCCS, across all 15 cell lines. This showed a greater than one order of magnitude range, from ca. 20 nM to 230 nM for PER2:CRY1 and CLOCK :CRY1 respectively (Figure 4A). Notably, affinity between PER2 and CRY1/2 was the strongest at 22.6 ± 1.3 nM (SD) and 30.1 ± 1.2 nM (SD) respectively. Additionally, PER2 showed an approximately 2-fold stronger affinity than PER1 did for both CRY proteins. PER2 readily formed homodimer complexes (K_D = 52.2 ± 5.6 nM, SD), with a moderate interaction also between CRY1 and CRY2 (K_D = 128.6 ± 6.7 nM, SD). In contrast, CLOCK and BMAL1 showed a relatively weak interaction in MEFs when compared to the circadian repressors PER and CRY, with a K_D of 82.9 ± 3.3 nM (SD). This accords with our previous measurements in NIH/3T3 cells (Koch et al., 2022). Remarkably, the affinity of CLOCK for BMAL1 was 3-fold greater than for the paralog BMAL2 (Figure 4A), which was predominately cytoplasmic (Figure 2A). Despite increased levels of expression in cell lines using lentiviral transduction, we observed generally a close concordance with the K_D values when measured from circadian-synchronised endogenously expressed proteins using primary fibroblasts derived from CRISPR-modified PER2xCRY1 and BMAL1xCRY1 mice (Figure 4A). Further, the affinity of these protein pairs changed significantly over

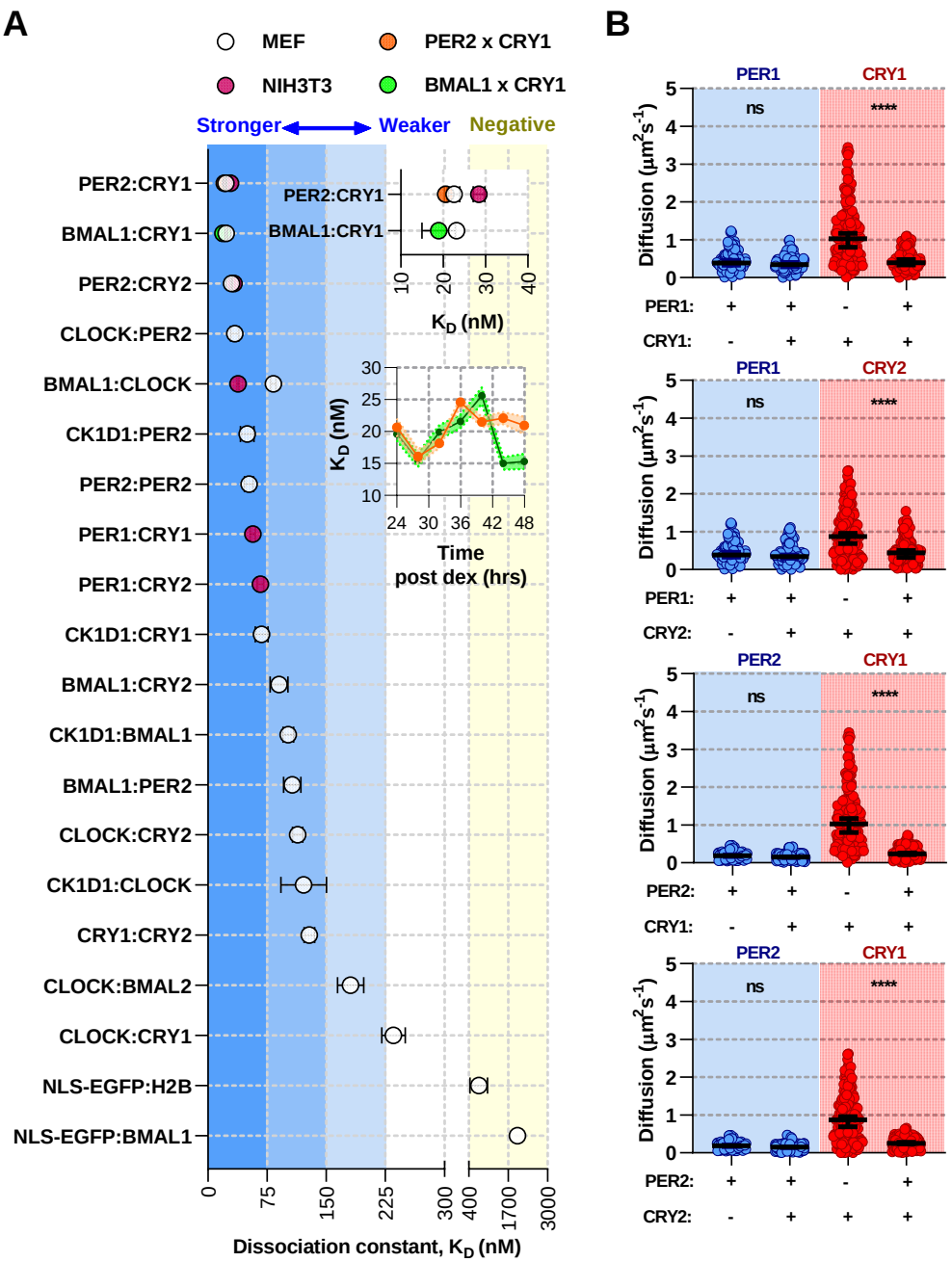

Figure 4: Live-cell interaction measurements of circadian protein pairs by FCCS (**A**) Live-cell FCCS measurements of interaction strength for pairs of fluorescent fusion proteins. Measurements were performed in four cell models; lentivirus transduced MEFs and NIH/3T3 cells (LV2) and dexamethasone synchronised primary cells lung fibroblasts isolated from BMAL1 x CRY1 fluorescent knock-in (KI) mice or skin fibroblasts isolated from PER2 x CRY1 fluorescent KI mice. Quantification of K_D for BMAL1 x CRY1 time-course and NIH/3T3 (PER2:CRY1) are a re-analysis of previous data in Koch et al. (2022). (**B**) Quantification of PER1/2 and CRY1/2 diffusion rates from FCCS measurements in NIH/3T3 fibroblasts. Statistical analysis was carried out using Kruskal-Wallis (values are denoted as $p < 0.001$ *** and $p < 0.0001$ ****).

the circadian cycle, and with near identical temporal pattern for PER2/CRY1 (measured here), and BMAL1/CRY1 (Koch et al., 2022).

In addition to binding strength (K_D), we assessed whether

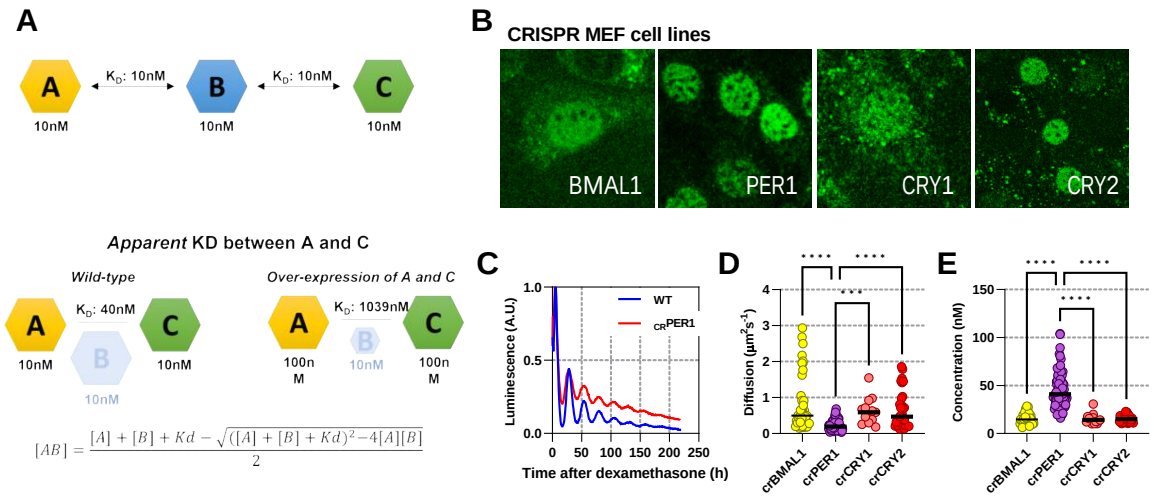

Figure 5: Framework to determine indirect and direct interactions using quantification of endogenous proteins. (**A**) Concept for calculating expected K_D from a series of paired interactions assuming that A and C do not interact. In this example, when A, B, and C are at comparable concentration the apparent K_D would be 40 nM, which is reduced to 1039 nM when proteins are over-expressed 10-fold as the intermediate partner, B, is limited. (**B**) Confocal microscopy images of polyclonal MEF-PER2::LUC cells genetically modified using CRISPR to express circadian target proteins C-terminally fused to EGFP. (**C**) To check for normal circadian rhythms after insertion of PER1::EGFP, PER2::LUC bioluminescence of clonal cells were measured following synchronisation by dexamethasone shock against unmodified WT MEF-PER2::LUC cells. FCCS was performed on nuclear fluorescence of all CRISPR modified cells, determining diffusion (**D**) and (**E**) concentration of EGFP labelled molecules. Statistical analysis was carried out using Kruskal-Wallis (values are denoted as $p < 0.001$ *** and $p < 0.0001$ ****).

protein-protein interactions may relate to changes in protein mobility, since larger complexes move slower (Konopka et al., 2006). As we observed changes to the fraction of CRY1 in complex with PER2 (Figure 1G,H), we hypothesised that the cycling in the diffusion rate of endogenous CRY1 (Figure 1F) could be caused by rhythmic binding with PER proteins. We assessed diffusion rates of CRY1/2 upon co-expression with PER1/2 in NIH/3T3 cells (Figure 4B), with MEF diffusion data for PER2 with CRY1 or CRY2 shown in Figure supplement 5. PER2 protein exhibited remarkably low diffusion rates, compatible with proteins with either high intrinsic disorder (Pelham et al., 2020) or as components in larger protein complexes (Aryal et al., 2017), nor was this altered by additional co-transduction with either CRY1 or CRY2 (Figure 4B). In contrast, the mobility of both CRY proteins was significantly affected by the addition of PER proteins, indicating CRY1/2 proteins are incorporated into slow moving complexes with PER1/2 (Figure 4B). Similarly, we observed a decrease in protein mobility for both BMAL1 and CLOCK when co-expressed (Figure supplement 5), compatible with a model involving direct DNA binding (Koch et al., 2022). Thus, we find high affinity interactions between circadian clock-regulating proteins impact on their mobility in the cell.

Determination of direct and indirect interaction topology from observed dissociation constants

Proteins can interact with one another directly or through intermediate partners, such as scaffold proteins, to form larger complexes. These intermediate partners can facilitate the formation of protein complexes, as well as modulate the activity and function of the interacting proteins. Here, we have quantified interactions between pairs of over-expressed proteins, but this fails to detect possible roles of additional proteins. These unseen partners could be involved in multivalent interactions, which would allow them to bind multiple proteins simultaneously, altering the strength of binding between the original partners. To understand better the role of these facilitating intermediate partners, we developed a modelling-based approach to explore the formation of three-protein complexes. In this framework we compared a K_D derived from a model of two proteins being brought together by a third protein to the observed K_D measurements (Figure 5A). If the predicted K_D arising from a new three-way interaction accords well with the observed measurements then the probability of a 3-way interaction is high. In contrast, if an additional partner protein fails to match measured K_D values, we conclude that the original two partner model is most likely.

In this framework, we require measurements of the concentrations and pairwise K_D between each of the three pro-

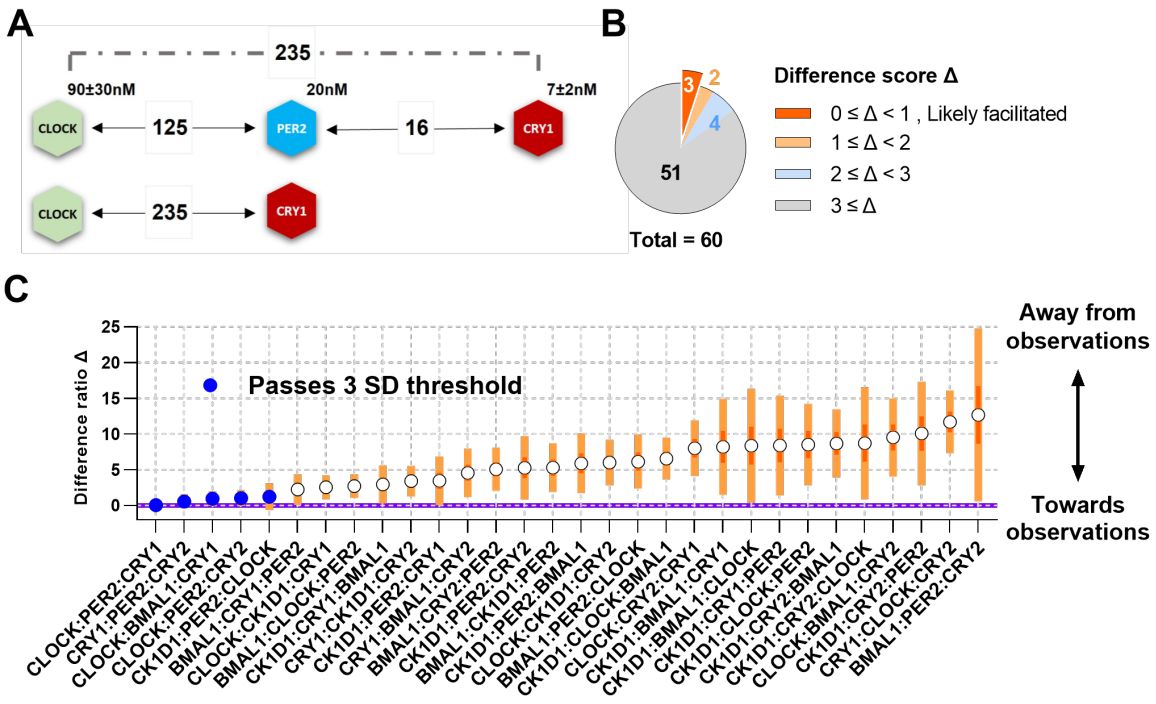

Figure 6: Calculating likelihood of indirect versus direct interactions from observed data. (**A**) A schematic of comparing a direct binding model versus a facilitated binding arrangement. If the calculated facilitated K_D matches the observed K_D closely then this arrangement is plausible. Example shows proposed models of direct and indirect interactions between CLOCK and CRY1. Indirect interactions occur via an intermediate partner, here shown as PER2. (**B**) Most K_D values arising from facilitated arrangements did not match the observed K_D, as assessed by the difference score Δ (see materials and methods). This was calculated assuming an indirect interaction model simulated using coupled ordinary differential equations (ODEs), an apparent K_D between a protein pair, A & C, was found for a middle partner, B, that facilitates the interaction. All concentration and K_D values were taken from experiments. (**C**) 60 three-protein arrangements were calculated to derive a difference ratio Δ between apparent K_D and experimentally observed K_D. Values closer to 0 indicate calculated K_D closely matches the experimental observed K_D. Chart shows data binned by difference ratio score Δ on the left- and right-hand side. Difference ratio values are plotted with 1 standard deviation (light orange) and 3 standard deviations (dark orange). Highly likely arrangements are highlighted in blue and are defined as calculated K_D values being within 3 standard deviations of the measured value.

teins. Our FCCS data provided the K_D values and concentrations of over-expressed partners required for these calculations (Figure 4A). However, in the model the additional partner protein would be expressed at endogenous concentrations. We therefore next set out to assess the behaviour of the endogenous proteins. We used CRISPR-mediated genetic editing to modify MEF-PER2::LUC cells to express either BMAL1, PER1, CRY1 or CRY2 fused with EGFP (Figure 5B). These modifications did not impact on molecular circadian oscillations, as assessed using PER2::LUC oscillations (Figure 5C). The diffusion rates for endogenous-tagged proteins were very similar to those in over-expressed lentiviral-transduced MEF cell lines (Figure supplement 5, Figure 5D,E). Our measures of endogenous concentration of circadian proteins in MEFs were a bit lower than the primary fibroblasts derived from genetically modified mice (compare Figure 5E vs Figure 1E, Figure supplement 2). Thus, protein behaviour as assessed by diffusion was consistent across genetically modified mice, CRISPR modified cell lines, and in lentiviral over-expressed proteins in MEF cells. In the case of PER2, genomic editing and quantification was complicated due to the insertion of luciferase to one PER2 allele, so to address this we used measures of endogenous PER2 protein concentration from fibroblasts derived from PER2::Venus mice. Hence, in the model, we used data for endogenous protein concentrations from the PER2 x CRY1 fibroblasts (Figure supplement 2F).

In order to develop a 3-way binding model (Figure 5A), an ordinary differential equation (ODE) was used to simulate the binding of proteins A and B through an intermediary protein C (Figure 6A), using our FCCS measurements for K_D and endogenous concentration values (Figure 5E). In total 60 protein combinations were used, modelling all possible non-repetitive three-way permutations. A difference

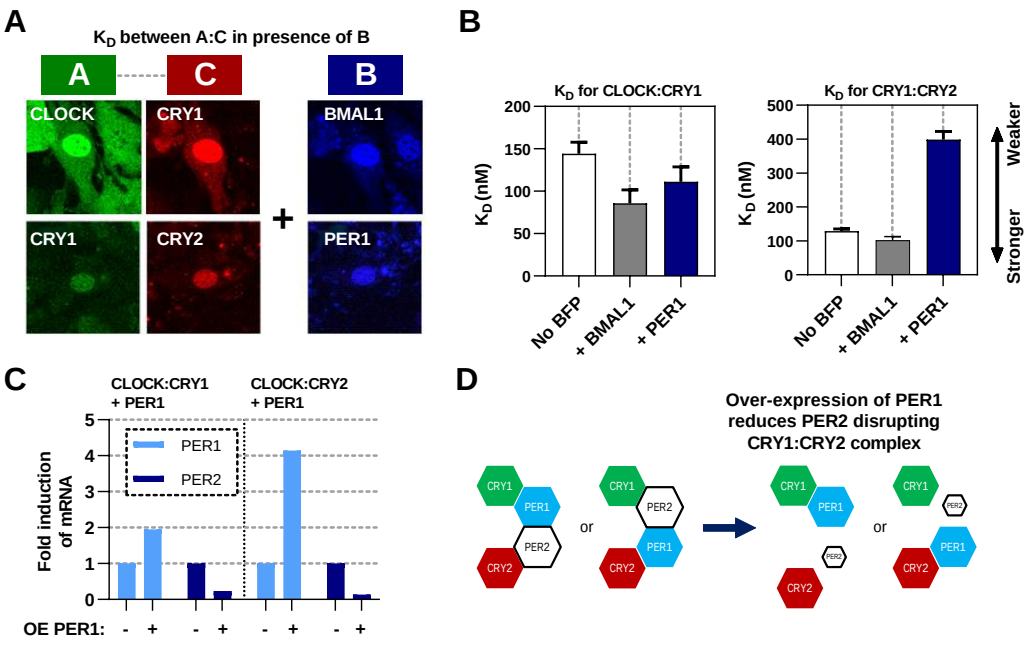

Figure 7: Three protein imaging reveals over-expression of PER1 can disrupt circadian macromolecular complex. (**A**) Confocal images of MEFs expressing EGFP and tagRFP fluorescent fusion proteins transduced to express an additional third blue tagBFP fusion protein (LV3 cells). (**B**) FCCS measurements of K_D between green and red proteins, in the presence or absence of a third blue protein. (**C**) RT-PCR of LV3 cells shows over-expression of PER1 reduces PER2 expression. (**D**) Cartoon of proposed tetramer model of PER:CRY in which loss of PER2 reduces measured indirect interaction between CRY1::EGFP and CRY2::tagRFP.

score, Δ, was used to rank the most likely indirect interactions, such that a score of 0 indicates that the calculated K_D arising from the indirect arrangement matches the experimentally observed K_D between those proteins A and C. A three-standard deviation (SD) uncertainty on the difference score, Δ, was propagated from the error on K_D values and concentrations. From these possible 60 combinations, only 5 had predicted Δ values within 3SD of 0. Strikingly, this suggests that the majority of protein-protein interactions are best defined as a pair-wise model, and without the need to propose involvement of an additional partner (Figure 6B). Nevertheless, these data support previous work showing a role for BMAL1 in enhancing CLOCK:CRY1 interactions, and for PER2 for in CK1δ:CLOCK interactions (Michael et al., 2017; An et al., 2022). Our data also show that CRY1:CRY2, and CLOCK:CRY2 depend upon additional partner proteins for complex formation (Figure 6C). Notably, out of the five triplet proteins here, PER2 was the key missing factor in four and BMAL1 in one (Figure 6C).

This modelling approach predicts possible three-way facilitated binding arrangements based on direct measures of the behaviour of two of the three components. In order to directly validate the model, we used our genetically modified LV2 MEF cells, expressing either CRY1::EGFP/CRY2::tagRFP or CLOCK::EGFP/CRY1::tagRFP, and then further transduced these cells to express a third protein, either BMAL1 or PER1, labelled with the blue fluorescent protein tagBFP. In this way, we were able to study the behaviour of proteins using 3 different fluorophores at different wavelengths denoted as LV3 cells (Figure 7A). From this we then measured K_D values between CLOCK and CRY1, finding the interaction to strengthen upon addition of PER1 or BMAL1 (Figure 7B). Under the same conditions, the affinity for CRY1 and CRY2 slightly increased, whereas added PER1 diminished the CRY1:CRY2 interaction complex by 4-fold (Figure 7B). Despite CRY1:PER2:CRY2 appearing at the top of predicted facilitated interactions in Figure 6C, the addition of PER1 reduced the formation of complexes incorporating CRY1 and CRY2. This suggests that any binding surface between CRY1 and CRY2 is interrupted by additional PER1. Thus, either PER1 may operate as a weaker partner than PER2 for CRY1:CRY2 complexes or that CRY1:CRY2 is brought together in four-way arrangements involving antagonistic PER proteins (Figure 7D). This is supported by RT-PCR, showing that overexpression of PER1 protein had significantly reduced endogenous Per2 mRNA (Figure 7C). Therefore, our data suggests that PER proteins are central mediators of macromolecular organisation of larger circadian complexes, with PER2 acting in a different roles to that of PER1.

Post-translational rhythms in CRY are generated by high affinity interactions that balance stoichiometry

The formation of protein complexes through protein-protein interactions can result in enhanced stability compared to the individual monomeric protein constituents. In our experiments using double-labelled, constitutively over-expressed LV2 MEFs, we observed significant correlations of protein abundance between several pairs of proteins. PER2 concentration was the strongest predictor of both CRY1 and CRY2 concentrations closely followed by BMAL1 (Figure 8A). Other pairwise protein combinations, including non-correlated controls, can be found in the full data set in Figure supplement 6. We investigated the correlations between green, red, and blue protein fluorescence in our triple labelled CLOCK:CRY1:(PER1 or BMAL1) and CLOCK:CRY2:(PER1 or BMAL1) LV3 lines. Upon co-expression with either PER1::tagBFP or BMAL1::tagBFP, red CRY1 fluorescence was markedly increased (Figure 8B). This was not a generalised responses to transduction, as we did not observe this effect with cells that expressed H2B::tagBFP (Figure 8B). As we have previously shown, BMAL1 re-localised CLOCK into the nucleus (Figure 8B) (Koch et al., 2022). Elevated PER1 protein strongly increased both CRY1 and CRY2 protein concentrations, regardless of CLOCK protein copy-numbers (Figure 8C), contrasting with the strong three-way correlation between BMAL1, CLOCK, and CRY1/2 expression (Figure 8D). As can be seen in Figure 8E, elevated PER1 and BMAL1 protein increased CRY1/2 concentrations to a near identical extent.

PER2 has been shown to bind and stabilise CRY2 protein (Schmalen et al., 2014; Nangle et al., 2014; Xing et al., 2013). If CRY protein stability depends on PER2 through high affinity interactions, it would be expected that CRY may oscillate without the requirement for rhythmic expression from its promoter. To test this, we performed simultaneous time-lapse measurements of fluorescence and bioluminescence following circadian synchronisation with DEX in our single labelled constitutively expressed CRY2::EGFP LV1 MEFs. This showed that CRY2 was both rhythmic and phase-lagged when compared with PER2 bioluminescence (Figure 8F). This phase-lag is consistent with our earlier published data derived from organotypic SCN slices of PER2::Venus and CRY1::mRuby3 mice (Smyllie et al., 2022) (Figure 8G). Although CRY1 may behave differently to CRY1. Thus, we propose two mechanisms driving rhythmic production of CRY, arising from transcription (Edwards et al., 2016) and as shown here, PER-mediated stabilisation of CRY through high-affinity interactions (Figure 8H).

Discussion

In this study, we aimed to shed light on the critical role of protein-protein interactions and multimeric complex formation in the functioning of the mammalian circadian clock. To this end, we introduced a novel approach using Fluorescence Cross-Correlation Spectroscopy (FCCS) to analyse the interactions of full-length proteins in their native cellular environment. Our approach allowed us to maintain normal and post-translational modifications and enabled us to assess protein localization, which are often lost in in-vitro studies. Our results revealed significant changes in protein localization upon co-expression with partner proteins PER-CRY and CLOCK-BMAL1. The data obtained from this study showed that the circadian clock operates primarily through pair-wise protein-protein interactions without the need for additional third partners. Our findings also revealed that the lifetime of CRY is dependent on PER2 and that a circadian clock protein (CRY2) can oscillate without the requirement for rhythmic transcriptional control. Our novel approach to dissect protein-protein interactions using FCCS has provided new insights into the functioning of the mammalian circadian clock, highlighting the importance of protein-protein interactions in regulating circadian rhythmicity. The use of live-cell methodologies in this study has offered a fresh perspective on the formation of protein complexes over time and space and has also opened the possibility of using this quantitative approach to explore other biological systems.

A growing model of the circadian clock proposes that the negative feedback on CLOCK:BMAL1 transcription factors, responsible for driving circadian rhythms, occurs via two modes of repression (Koike et al., 2012; Cao et al., 2021, 2023; Koch et al., 2022). The first mode, called "displacement", occurs when the PER:CRY complex translocates CK1δ/ϵ into the nucleus to phosphorylate CLOCK:BMAL1 dimers and remove them from DNA. The second mode, called "blocking", takes place after PER degradation and involves CRY1 binding to CLOCK:BMAL1 and blocking the recruitment of transcriptional co-factors to the c-terminal transactivation domain of BMAL1 (Xu et al., 2015). Hence, this model places the relative abundance of protein complexes through time and space as a key mechanism for the regulation of circadian transcriptional activity. In vitro studies have found a vast array of the repressive CK1-PER-CRY ternary complexes are made up of differing ratios of protein paralogs, whose, differential functions are hard to tease apart (Aryal et al., 2017; Cao et al., 2021, 2023). The differential identities of this key complex are due to the many possible combinations of PER1-3 and CRY1/2 proteins which is modulated by differences in protein expression and divergent affinities of these paralogs for their key partner proteins. In our study, we find that both CRY2 and PER1 are weaker interactors for CLOCK, BMAL1, and CK1δ than CRY1 and PER2 respectively. Furthermore, PER1 is expressed 2-fold over PER2 on average confirming the findings by (Zheng et al., 2001), whereas CRY1 and CRY2 protein abundances are equal. Taken together, this shows that both protein abundance and protein-protein affinity balance and tune the stoichiometric identity of key CK1-

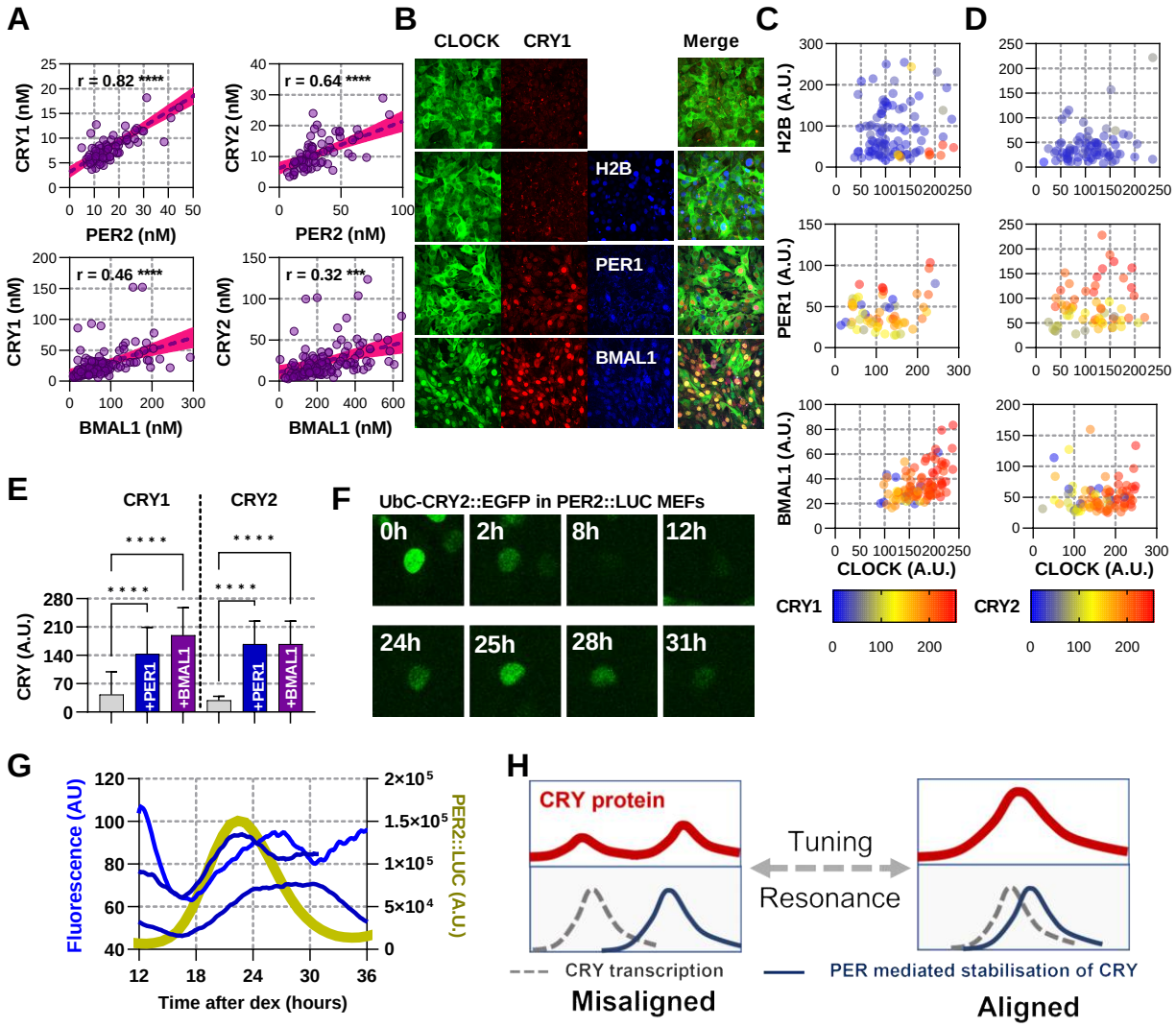

Figure 8: High affinity interactions and stochiometric balancing ensure post-translational rhythms of CRY. (**A**) Analysis of protein correlations from FCCS measurements. LV2 MEF cells used to measure interactions, were re-analysed for nuclear protein concentrations, looking for correlations between the two proteins despite coming from separate transductions and expressed from the constitutive ubiquitin ligase C promoter. Spearman correlation score is shown alongside linear regression. (**B**) Cells expressing CLOCK and CRY as green EGFP and red tagRFP fluorescent fusion proteins were transduced to express a third blue fluorescent fusion protein (referred to as LV3 cells). Multivariate correlation of fluorescence between EGFP, tagBFP, and tagRFP, with tagRFP levels indicated by colour intensity for (**C**) CRY1::tagRFP (**D**) and CRY2::tagRFP.(**E**) Quantification of CRY fluorescence after transduction with third blue fusion protein. (**F**) Timelapse confocal images of MEF-PER2::LUC LV1 cells constitutively expressing CRY2::EGFP following dexamethasone shock circadian synchronisation. (**G**) Quantification of single cell nuclear fluorescence for three CRY2::EGFP cells (blue lines) shown alongside superimposed separate recordings of PER2::LUC bioluminescence (green line) also made in CRY2::EGFP cells. (**H**) Proposed model for PER protein accumulation and high affinity PER:CRY interactions results in stabilisation of CRY proteins providing a tuning mechanism for the clock. Statistical analysis was carried out using Kruskal-Wallis (values are denoted as $p < 0.001$ *** and $p < 0.0001$ ****).

PER-CRY repressive complexes over circadian cycles.

Our results show that both CRYs and PERs are capable of tightly associating with CLOCK:BMAL1. However, it is not yet clear whether PER1/2 is capable of directly binding CLOCK or BMAL1, or if CRY1/2 brings PER1/2 into multi-meric complexes with CLOCK and BMAL1. The lack of in vitro evidence for direct binding of PER1/2 with CLOCK:BMAL1 and strong interaction between CRY1-PER2, as well as CRY1-BMAL1, points to the latter model. Overall, These interactions are remarkably strong and occur in the low nanomolar range, comparable with pharmacological levels of affinity (Ma et al., 2018). As such, the majority of these

proteins are bound in complex, rarely existing as monomers, confirming in vitro observations (Aryal et al., 2017; Cao et al., 2021, 2023). Strikingly, the interaction of PER2 with CRY1 was found to be almost indistinguishable in strength when compared with our previously characterised BMAL1-CRY1 interaction over 24 hours (Koch et al., 2022), with the ratio of these proteins in complex oscillating in phase up to the point at which CRY1 "blocking" repression began to dominate. In accordance with this, the expression of CRY1 phase-lags that of PER2 (Smyllie et al., 2022). This provides the first direct evidence in living cells for the operation of the temporal shift between blocking and displacement modes of CLOCK:BMAL1 repression.

We find that, with the exception of CLOCK-BMAL1 interactions, circadian proteins form tight association with one another, and this is conserved irrespective of cell-type or the abundance of protein. This accords with the fact that the spatiotemporal regulation of CK1δ/ϵ and PER (FRQ in Neurosprora) complex formation is conserved across eukaryotes (Wong and O'Neill, 2018). Additionally, evidence shows that the transactivation domain (TAD) of BMAL1, which is bound by CRY1, is highly conserved amongst vertebrates and also in insects with vertebrate-like clocks (Yuan et al., 2007). Similarly, PAS domains of PER proteins share significant similarity in *Drosophila* and mammalian clocks and basic-Helix-Loop-Helix (bHLH)-PAS transcription factors (CYC, BMAL, CLK, and NPAS2) (Huang et al., 1993). Together with our data, this suggests that protein-protein interactions are an evolutionarily conserved mechanism for circadian rhythm generation, with differences in protein components and expression levels providing a tuning mechanism between species and cell-types.

Prior studies have shown a high degree of variance in BMAL1 expression within cultures and between cell-types (Yang et al., 2020; Koch et al., 2022). BMAL1 is well established as a key driver of the circadian clock, with its deletion abolishing circadian rhythms (Haque et al., 2019). Furthermore, when BMAL1 is expressed constitutively (i.e. non-rhythmically), normal circadian rhythms may still occur (Preitner et al., 2002). Taken with our finding that BMAL1 plays an important trafficking role allowing nuclear entry by CLOCK, and that the interaction with its repressors generates oscillations, BMAL1 operates as a key nodal point of the circadian system.

PERs possess distinct binding sites for CK1δ/ϵ and CRY1-/2 to act as a scaffold for crucial CK1:PER:CRY ternary complexes (Vielhaber et al., 2000; Cao et al., 2021). One of these sites is the PAS domain of PER, which permits homodimerization as well as heterodimerisation between PER1 and PER2 (Yagita et al., 2000; Hennig et al., 2009; Kucera et al., 2012). Removal of the core PAS-B in the Per2^{Brdm} mutant leads to loss of circadian function (Zheng et al., 1999). We report here the first direct evidence in the mammalian clock that PER2:PER2 homodimers are readily detectable, which possibly supports the formation of CRY1:CRY2 complexes. Given that this interaction is ablated upon over-expression of PER1 and reduction of PER2 we conclude that PER:PER homo- and hetero-dimers may likely incorporate one or more CRY proteins into a tetramer complex. In this model PER2 homodimerization would enable macromolecular assembly, exemplified by CRY1-CRY2 binding.

In this study, we find that protein abundance is regulated by complexing with partner proteins. Simultaneous over-expression demonstrates that levels of CRY1/2 protein are correlated with BMAL1 and PER2, proteins despite their constitutive expression from UbC promoters. These strong correlations provide evidence of post-translational regulation of protein stability. This confirms reports of PER2 stabilising CRY2 by blocking degradation of CRY2 by FBXL3 (Nangle et al., 2014; Parlak et al., 2022). Stabilisation of CRY1/2 when co-expressed alongside BMAL1, with or without extra CLOCK, may represent a novel mechanism of BMAL1 directly protecting CRY1/2 from degradation. We show how PER regulates both localization, diffusion, and stability of CRY through a very high affinity interaction, which is able to confer rhythmicity to CRY even in the absence of cycling Cry mRNA. This defines a PER2 mediated post-translational mechanism for generating rhythmic CRY. In this model, CRY proteins would have two sources of rhythmicity, which can be tuned to one another: regulated transcription and regulated degradation. Both of these work together to confer rhythmicity to CRY, and may act in phase (constructive interference) or out of phase (destructive interference) to tune circadian rhythms.

Since the discovery of post-translational circadian rhythms in phosphorylation of KaiC in *Cyanobacteria* (Nakajima et al., 2005), a number of studies have explored whether a classical transcription-translation feedback loop (TTFL) or post-translational oscillator (PTO) forms the fundamental core circadian mechanism in eukaryotes (Wong and O'Neill, 2018; Partch, 2020; Crosby and Partch, 2020). In this study, we find that the interplay between post-translational mechanisms and expression levels set by transcription are closely related and operate together reciprocally to generate robust circadian rhythms.

Materials and Methods

Plasmids

A collection of lentivirus transfer plasmids was cloned allowing constitutive expression of fluorescent fusion proteins from the Ubiquitin ligase C promoter. Plasmids were cloned using the gateway cloning system as previously described. In brief, an initial 'entry' vector was cloned containing the murine coding sequences for: Csnk1d (NM_139059.3), and Cry2 (NM_009963.4) and Per1 (NM_011065.5) adding to an existing library of several entry vectors encoding Bmal1 (NM_007489.4), Clock (NM_007715.6), Cry1 (NM_007771.3)

and Per2 (NM_011066.3). Destination vectors were then produced by recombining an entry vector with a target destination vector encoding a fluorescent protein, either EGFP, tagRFP or tagBFP as described in (Bagnall et al., 2015). Both the pLNT-H2B-RFP and NLS::EGFP vectors were published previously (Bagnall et al., 2015; Koch et al., 2022).

Animal lines and isolated primary cells

Mice were maintained in the University of Manchester Biological Services Facility. Skin fibroblasts were isolated from a previously described C57BL/6 mouse knock-in line expressing PER2::Venus and CRY1::mRuby3 (Koch et al., 2022; Smyllie et al., 2022). In brief, ear punches were taken, homogenised by scalpel blade, and resuspended in DMEM (SIGMA; Cat no. D6429) supplemented with 10% fetal bovine serum (HyClone; Cat No. 12379802) and penicillin/streptomycin (10 U/ml). Cultures were then left to grow over the course of 2 weeks replacing media every 3-4 days, with additional care to agitate or remove large tissue pieces. To prepare cells for experimentation, cells were detached from culture vessels using trypsin and plated at 2×10^5 cells per 35 mm dish.

Maintenance and generation of transduced cell lines

Mouse embryonic fibroblasts containing the heterozygous PER2::LUC fusion were cultured in DMEM/F12 (Gibco; Cat. No. 11320033) supplemented with 10% bovine fetal calf serum. MEFs were derived from previously reported PER2::LUC mouse (Yoo et al., 2004, 2017). Cells were seeded into T25 flasks and sub-cultured every 3-4 days. Cells were tested for the absence of mycoplasma using MycoAlert mycoplasma detection kit (Cat. No. LT07-418). MEF cells were used to produce fluorescent labelled cells via lentivirus transduction or CRISPR modification. Lentivirus was produced as previously reported (Bagnall et al., 2015), transfecting a third-generation plasmid set into HEK293T cells and harvesting virus 3 days later by ultracentrifugation. Virus was then either immediately used to transduce by directly pipetting on to 150,000 MEFs in 2ml culture media or frozen at -80°C for later use. Transduction of the first fluorescent protein was then confirmed by confocal microscopy. Cells transduced once to express a single fluorescent protein were termed LV1. A second sequential transduction was used on cells after a minimum of 2 weeks to introduce a further fluorescent protein (termed LV2 cells) and this was continued to a third transduction to produce green, red, and blue LV3 cells.

Generation of knock-in cell lines

We used CRISPR-Cas9 to tag endogenous genes with fluorescent fusion genes. A generic reporter vector was generated, comprising linker-EGFP-T2A-mScarletI-DestabilisationDomain-Loxp-pgk-Puro-LoxP. Homology arms (~ 800 bp) directed to the 3' end of each target gene was HiFi assembled (NEB) into the vector. This donor was then co-transfected with Cas9 and sgRNA (IDT) designed to generate a double strand break over the Stop codon of the target gene into MEF Per2-Luc cells. Knock-in cells were selected for by Puromycin selection (2.5 μ/ml; Gibco Cat No. A1113803).

Circadian synchronization

Circadian synchronisation of cells was achieved by stimulation with 200 nm dexamethasone (Sigma D4902) for 1 hr before PBS washes and then switched to fresh culture media. Cultures were passaged for biological replicates.

Confocal microscopy

Imaging was performed by confocal microscopy using a ZEISS LSM880 microscope equipped with equipped with a stage mounted incubator to maintain cells at 37°C in humidified 5% CO2. Fluorescence image capture was performed using ZEN 2.1 SP3 FP2 software and all images were made using a FLUAR 40 x NA 1.3 oil immersion objective. Samples were prepared by plating cells on to 35 mm glass bottomed imaging dishes (Greiner Bio-one) then imaging cells 24-48 hours later. A range of fluorescent proteins were imaged by different excitation lasers; 405 nm excitation was used for tagBFP, 488 nm for EGFP & Venus, and 561 nm for tagRFP and mRuby3. Array detectors were used to select and collect the appropriate emitted fluorescence. Analyses of nuclear/cytoplasmic localisation were made using cell tracker ver 0.1 to segment signals (Du et al., 2010).

Fluorescence Cross Correlation Spectroscopy

Images of cells by confocal microscopy were made as above and were used to select a point to measure in the nucleus of fluorescent labelled cells. FCCS measurements were made using 40x objective set to 1 airy unit after calibrating and adjusting pinhole through software, resulting in a approximate measurement volume of 0.722 fl and 1.10 fl for EGFP (488 nm excitation) and tagRFP (561 nm excitation) respectively. Measurements were performed for 20 seconds per nuclei, with care to minimize bleaching and maintain signal counts above 0.3 kHz. Auto- and cross-correlation data was then analysed using our previously reported FCS Analysis Pipeline in Python 3.9.7 described in Koch et al. (2022) (GitHub: github.com/LoudonLab/FcsAnalysisPipeline), to determine concentration and diffusion amongst other parameters. Several updates were made to the analysis script available as version 2 through GitHub. In brief, models described anomalous diffusion were fit to all correlation traces and interaction strength determined by robust linear fitting free and complexed concentrations. The approximate confocal volumes were calculated via the equation

$$V_{\text{eff}} = (2\pi)^{\frac{3}{2}} w_{xy}^2 w_z = (2\pi)^{\frac{3}{2}} \left(\frac{0.61\lambda}{\text{NA}}\right)^2 \left(\frac{2n\lambda}{\text{NA}^2}\right), \quad (1)$$

where w_{xy} and w_z is the beam width in the xy plane and z axis respectively, with NA as the numerical aperture (NA = 1.3 for our 40x objective), λ the wavelength of exciting laser and n the refractive index of the immersion oil ($n = 1.515$ in all experiments). The appropriate spectra were collected for each different fluorophore. Laser power was reduced to minimise photo-bleaching whilst maintaining counts per molecule greater than 0.3 kHz. Care was taken for Fluorescence cross-correlation spectroscopy (FCCS) measurements to avoid the green channel signal spilling up into the red channel causing false cross-correlation by reducing the laser power and observing the far-red part of the second channel. Control measurements were performed by selectively turning off either 488 or 561 nm lasers and tuning the red channel spectra until there was no cross-correlation due to spill-over. To calculate the disassociation constant K_D we plotted FCCS data according to the model

$$K_D = \frac{[A]_{\text{Free}}[B]_{\text{Free}}}{[\text{Complex}]} \quad (2)$$

where $[\cdot]$ denotes Molar concentration. This is in accordance with the same linear fit procedure as Sadaie et al. (2014). Concentrations of free fluorescent proteins were calculated from the green and red auto-correlations and the concentration of complex as

$$[X]_{\text{Free}} = [X]_{\text{Total}} - [\text{Complex}] \quad (3)$$

where $[X]_{\text{Total}}$ is the total concentration of the protein X as assessed from each auto-correlation. Linear fits to these data were achieved through robust quantile regression for an intercept of zero and taking the median slope. The slope and its standard deviation were calculated by the *QuantileRegressor* algorithm implemented in the Python package statsmodels (ver. 0.13.5) (Seabold and Perktold, 2010). Other standard numerical algorithms were performed using SciPy (ver. 1.9.3) and NumPy (ver. 1.24.1) (Virtanen et al., 2020; Harris et al., 2020). The percentages of the proteins A and B bound in complex was calculated from the ratio of total protein to complex as $A_{\text{Bound}} = [AB]_{\text{eq}}/[A]_{\text{Tot}}$ and $B_{\text{Bound}} = [AB]_{\text{eq}}/[B]_{\text{Tot}}$. The concentration $[AB]_{\text{eq}}$ was found using the formula

$$[AB]_{\text{eq}} = \frac{1}{2}\big([A]_{\text{Tot}} + [B]_{\text{Tot}} + K_D - \sqrt{([A]_{\text{Tot}} + [B]_{\text{Tot}} + K_d)^2 - 4[A]_{\text{Tot}}[B]_{\text{Tot}}}\big), \quad (4)$$

where $[A]_{\text{Tot}}$ and $[B]_{\text{Tot}}$ are the total concentrations of each protein, and K_D is the dissociation constant between them (Koch et al., 2022). The uncertainty on these bound fractions, A_{Bound} and B_{Bound}, was found by propagating the error on $[AB]_{\text{eq}}$, $[A]_{\text{Tot}}$, and $[B]_{\text{Tot}}$ assuming Gaussian error.

PMT recordings

MEFs expressing PER2::luciferase were plated into 35 mm plastic tissue culture dishes (Corning). D-luciferin (1 mM) was supplemented into the media 4–24 hr prior to recordings and dishes were sealed with parafilm. Bioluminescence was then recorded every minute by photomultiplier tubes (PMTs; Hamatasu) housed in an enclosed incubator at 37°C and without CO_2 as described previously (Loudon et al. 2007).

RT-PCR

Cells were plated in 35-mm tissue culture dishes (2x105 cells per dish) before isolation the following day. Total RNA was extracted from the cells with the ReliaPrepTM RNA miniprep kit (Promega) in accordance with manufacturer's instructions. RNA-to-cDNA conversion was achieved using High-Capacity RNA-to-cDNA™ Kit (Applied Biosystems). Real-time qPCR was performed on the StepOnePlus™ Real-Time PCR System using the KAPA SYBR® FAST qPCR Master Mix (2X) Kit (Sigma-aldrich). Relative quantification of mRNA abundance was conducted by normalization to Cyclophillin A mRNA. The primers used are specified in Table 1 Oligonucleotides sequences used in this study are as follows (5' to 3'): PER etc.

Gene	Primer sequence
Per1 Forward	AGGTGGCTTTCGTGTTGG
Per1 Reverse	CAATCGATGGATCTGCTCTGAG
Per2 Forward	CCTACAGCATGGAGCAGGTTGA
Per2 Reverse	TTCCCAGAAACCAGGGACACA
Bmal1 Forward	GGCTGTTCAGCACATGAAAAC
Bmal1 Reverse	GCTGCCCTGAGAATTAGGTGTT
Cry1 Forward	GTGGATCAGCTGGGAAGAAG
Cry1 Reverse	CACAGGGCAGTAGCAGTGAA
Cry2 Forward	CCCACGGCCCATCGT
Cry2 Reverse	TGCTTCATTCGTTCAATGTTGAG

Table 1: Sequences of primers used in RT-PCR

Statistical analysis

The data presented was analysed GraphPad Prism version 9.5 for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com. An appropriate normality test was performed to assess whether the data fit a Gaussian distribution. Non-parametric and parametric tests were used as appropriate and indicated in figure legends.

Mathematical modelling of facilitated interactions

We used ODEs to model indirect binding of species A and C, facilitated by an intermediate protein B, to create an

observable complex, A:B:C. This incorporates A, B, and C all within the same trimer protein complex. In this model A binds with B into the dimer A:B, and B binds with C to make dimer B:C. Subsequently, both A:B and B:C can bind with C and A respectively to create the final complex A:B:C. We make use of the definition of K_D between species A and B

$$K_D(A,B) = \frac{k_d}{k_a} = \frac{[A][B]}{[AB]} \tag{5}$$

to equate the backwards rate of reaction as $k_d = k_a K_D(A,B)$ (Zhou and Bates, 2013; Sadaie et al., 2014). Here, concentrations of each species are denoted at $[\cdot]$, i.e. $[A]$ represents the concentration of A. As we are only concerned with the equilibrium concentrations of monomers and complexes, we set all association rates to be equal and place all the differences in interaction rates in the K_D between each pairing. This gives us the set of ODEs

$$\begin{aligned}
\frac{d[A]}{dt} &= k_a(-[A][B] + K_D(A,B)[A{:}B] \\
&\quad - [A][B{:}C] + K_D(A,B)[A{:}B{:}C]), \\
\frac{d[B]}{dt} &= k_a(-[A][B] + K_D(A,B)[A{:}B] \\
&\quad - [B][C] + K_D(B,C)[B{:}C]), \\
\frac{d[C]}{dt} &= k_a(-[B][C] + K_D(B,C)[B{:}C] \\
&\quad - [A{:}B][C] + K_D(B,C)[A{:}B{:}C]), \\
\frac{d[A{:}B]}{dt} &= k_a(+[A][B] - K_D(A,B)[A{:}B] \\
&\quad - [A{:}B][C] + K_D(B,C)[A{:}B{:}C]), \\
\frac{d[B{:}C]}{dt} &= k_a(+[B][C] - K_D(B,C)[B{:}C] \\
&\quad - [A][B{:}C] + K_D(A,B)[A{:}B{:}C]), \\
\frac{d[A{:}B{:}C]}{dt} &= k_a(+[A{:}B][C] - K_D(B,C)[A{:}B{:}C] \\
&\quad + [A][B{:}C] - K_D(A,B)[A{:}B{:}C]),
\end{aligned} \tag{6}$$

where we have made use of the definitions $K_D(A,B{:}C) = K_D(A,B)$ and $K_D(A{:}B,C) = K_D(B,C)$. This is because B facilities the interaction between A and C as there is no interaction between A and C, such that $K_D(A,C) = 0$. We numerically solve these equations from an initial time $t = 0$ until an equilibrium is established using the SciPy (ver. 1.9.3) *odeint* integrator in Python (ver. 3.9.2) (Virtanen et al., 2020). Initial concentrations of all complexes A:B, B:C, and A:B:C are set to zero and concentrations monomers of monomers are set as total concentrations of each protein from our data. For each arrangement of A, B, and C we have used concentrations derived from FCS of LV2 lines for proteins A and C, and endogenous FCS measurements from CRISPR engineered MEFs for B. For example in the arrangement CLOCK:BMAL1:CRY1, CLOCK and CRY1 concentrations and K_D values (Figure 4A and Figure supplement 6) are taken from LV2 measurements whereas BMAL1 was set at the concentrations measured for CRISPR MEFs, see Figure 5E. K_D values were similarly taken from FCCS measurements made in over-expressed LV2 MEFs. In the case of arrangements involving PER2 as the mediator, B, no CRISPR MEF cell-line was available for FCS measurements, hence nuclear concentrations were determined from non-DEX asynchronous skin fibroblasts isolated from the PER2::Venus CRY1::mRuby3 double knock-in mouse. The difference score, Δ, which represents how close the calculated $K_D^{(\mathrm{calc})}$ is the the observed $K_D^{(\mathrm{obs})}$ is calculated as

$$\Delta = \frac{K_D^{(\mathrm{calc})}}{K_D^{(\mathrm{obs})}} - 1. \tag{7}$$

Acknowledgments

We are profoundly grateful to Dr. Jean-Michel Fustin for generously providing the PER2::LUC mouse embryonic fibroblast (MEF) cells for our study. We thank the Bioimaging core facility at the University of Manchester for all their assistance. We also thank the University of Manchester Flow Cytometry Core Facility for assistance with sorting CRISPR cells. Supported by the Biotechnology and Biological Sciences Research Council, UK (award BB/P017347/1 to MHH and ASIL), MRC core funding (MC_U105170643 to MHH), the National Institutes of Health (award GM107069 and GM141849 to CLP). ASIL is a Wellcome Investigator (Grant 107851/Z/15/Z). AAK is supported by a Wellcome 4 year PhD studentship (216416/Z/19/Z).

References

An Y, Yuan B, Xie P, Gu Y, Liu Z, Wang T, Li Z, Xu Y, Liu Y. Decoupling PER phosphorylation, stability and rhythmic expression from circadian clock function by abolishing PER-CK1 interaction. Nature Communications. 2022 Jul; 13(1):3991. https://doi.org/10.1038/s41467-022-31715-4, doi: 10.1038/s41467-022-31715-4.

Aryal RP, Kwak PB, Tamayo AG, Gebert M, Chiu PL, Walz T, Weitz CJ. Macromolecular Assemblies of the Mammalian Circadian Clock. Molecular Cell. 2017 Sep; 67(5):770–782.e6. https://doi.org/10.1016/j.molcel.2017.07.017, doi: 10.1016/j.molcel.2017.07.017, publisher: Elsevier.

Bacia K, Kim SA, Schwille P. Fluorescence cross-correlation spectroscopy in living cells. Nature Methods. 2006 Feb; 3(2):83–89. https://www.nature.com/articles/nmeth822, doi: 10.1038/nmeth822, number: 2 Publisher: Nature Publishing Group.

Bagnall J, Boddington C, Boyd J, Brignall R, Rowe W, Jones NA, Schmidt L, Spiller DG, White MRH, Paszek P. Quantitative dynamic imaging of immune cell signalling using lentiviral gene transfer. Integrative Biology. 2015 Jun; 7(6):713–725. https://doi.org/10.1039/c5ib00067j, doi: 10.1039/c5ib00067j, number: 6.

Buhr ED, Takahashi JS. Molecular components of the Mammalian circadian clock. Handb Exp Pharmacol. 2013; (217):3–27. https://www.ncbi.nlm.nih.gov/pubmed/23604473, doi: 10.1007/978-3-642-25950-0_1, type: Journal Article.

Buijs RM, la Fleur SE, Wortel J, van Heyningen C, Zuiddam L, Mettenleiter TC, Kalsbeek A, Nagai K, Niijima A. The suprachiasmatic nucleus balances sympathetic and parasympathetic output to peripheral organs through separate preautonomic neurons. Journal of Comparative Neurology. 2003 Sep; 464(1):36–48. https://doi.org/10.1002/cne.10765, doi: 10.1002/cne.10765, publisher: John Wiley & Sons, Ltd.

Cao X, Wang L, Selby CP, Lindsey-Boltz LA, Sancar A. Analysis of Mammalian Circadian Clock Protein Complexes Over a Circadian Cycle. Journal of Biological Chemistry. 2023 Jan; p. 102929. https://www.sciencedirect.com/science/article/pii/S0021925823000613, doi: 10.1016/j.jbc.2023.102929.

Cao X, Yang Y, Selby CP, Liu Z, Sancar A. Molecular mechanism of the repressive phase of the mammalian circadian clock. Proceedings of the National Academy of Sciences. 2021 Jan; 118(2):e2021174118. https://doi.org/10.1073/pnas.2021174118, doi: 10.1073/pnas.2021174118, publisher: Proceedings of the National Academy of Sciences.

Crosby P, Partch CL. New insights into non-transcriptional regulation of mammalian core clock proteins. Journal of Cell Science. 2020 Sep; 133(18):jcs241174. https://doi.org/10.1242/jcs.241174, doi: 10.1242/jcs.241174.

Du CJ, Marcello M, Spiller DG, White MRH, Bretschneider T. Interactive segmentation of clustered cells via geodesic commute distance and constrained density weighted Nyström method. Cytometry Part A. 2010 Dec; 77A(12):1137–1147. https://doi.org/10.1002/cyto.a.20993, doi: 10.1002/cyto.a.20993, publisher: John Wiley & Sons, Ltd.

Edwards MD, Brancaccio M, Chesham JE, Maywood ES, Hastings MH. Rhythmic expression of cryptochrome induces the circadian clock of arrhythmic suprachiasmatic nuclei through arginine vasopressin signaling. Proceedings of the National Academy of Sciences. 2016 Mar; 113(10):2732–2737. https://doi.org/10.1073/pnas.1519044113, doi: 10.1073/pnas.1519044113, publisher: Proceedings of the National Academy of Sciences.

Haque SN, Booreddy SR, Welsh DK. Effects of BMAL1 Manipulation on the Brain's Master Circadian Clock and Behavior. Yale J Biol Med. 2019 Jun; 92(2):251–258. https://www.ncbi.nlm.nih.gov/pubmed/31249486, place: United States.

Harris CR, Millman KJ, Walt SJvd, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, Kerkwijk MHv, Brett M, Haldane A, Río JFd, Wiebe M, Peterson P, Gérard-Marchant P, et al. Array programming with NumPy. Nature. 2020 Sep; 585(7825):357–362. https://doi.org/10.1038/s41586-020-2649-2, doi: 10.1038/s41586-020-2649-2, publisher: Springer Science and Business Media LLC.

Hennig S, Strauss HM, Vanselow K, Yildiz Schulze S, Arens J, Kramer A, Wolf E. Structural and Functional Analyses of PAS Domain Interactions of the Clock Proteins Drosophila PERIOD and Mouse PERIOD2. PLOS Biology. 2009 Apr; 7(4):e1000094. https://doi.org/10.1371/journal.pbio.1000094, doi: 10.1371/journal.pbio.1000094, publisher: Public Library of Science.

Hess ST, Webb WW. Focal volume optics and experimental artifacts in confocal fluorescence correlation spectroscopy. Biophysical Journal. 2002 Oct; 83(4):2300–2317. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1302318/, number: 4.

Huang ZJ, Edery I, Rosbash M. PAS is a dimerization domain common to Drosophila Period and several transcription factors. Nature. 1993 Jul; 364(6434):259–262. https://doi.org/10.1038/364259a0, doi: 10.1038/364259a0.

Koch AA, Bagnall JS, Smyllie NJ, Begley N, Adamson AD, Fribourgh JL, Spiller DG, Meng QJ, Partch CL, Strimmer K, House TA, Hastings MH, Loudon AS. Quantification of protein abundance and interaction defines a mechanism for operation of the circadian clock. eLife. 2022 Mar; 11:e73976. https://doi.org/10.7554/eLife.73976, doi: 10.7554/eLife.73976, publisher: eLife Sciences Publications, Ltd.

Koike N, Yoo SH, Huang HC, Kumar V, Lee C, Kim TK, Takahashi JS. Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals. Science. 2012; 338(6105):349–354. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3694775, doi: 10.1126/science.1226339, type: Journal Article.

Komatsubara AT, Goto Y, Kondo Y, Matsuda M, Aoki K. Single-cell quantification of the concentrations and dissociation constants of endogenous proteins. Journal of Biological Chemistry. 2019 Apr; 294(15):6062–6072. https://www.jbc.org/article/S0021-9258(20)36677-1/abstract, doi: 10.1074/jbc.RA119.007685, number: 15 Publisher: Elsevier.

Konopka MC, Shkel IA, Cayley S, Record MT, Weisshaar JC. Crowding and Confinement Effects on Protein Diffusion In Vivo. Journal of Bacteriology. 2006 Sep; 188(17):6115–6123. https://doi.org/10.1128/JB.01982-05, doi: 10.1128/JB.01982-05, publisher: American Society for Microbiology.

Konopka RJ, Benzer S. Clock Mutants of Drosophila melanogaster. Proceedings of the National Academy of Sciences. 1971 Sep; 68(9):2112–2116. https://doi.org/10.1073/pnas.68.9.2112, doi: 10.1073/pnas.68.9.2112, publisher: Proceedings of the National Academy of Sciences.

Kucera N, Schmalen I, Hennig S, Öllinger R, Strauss HM, Grudziecki A, Wieczorek C, Kramer A, Wolf E. Unwinding the differences of the mammalian PERIOD clock proteins from crystal structure to cellular function. Proceedings of the National Academy of Sciences. 2012 Feb; 109(9):3311–3316. https://doi.org/10.1073/pnas.1113280109, doi: 10.1073/pnas.1113280109, publisher: Proceedings of the National Academy of Sciences.

Lambert TJ. FPbase: a community-editable fluorescent protein database. Nature Methods. 2019 Apr; 16(4):277–278. https://www.nature.com/articles/s41592-019-0352-8, doi: 10.1038/s41592-019-0352-8, number: 4 Publisher: Nature Publishing Group.

Ma W, Yang L, He L. Overview of the detection methods for equilibrium dissociation constant KD of drug-receptor interaction. Journal of Pharmaceutical Analysis. 2018 Jun; 8(3):147–152. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6004624/, doi: 10.1016/j.jpha.2018.05.001, number: 3.

Meyer-Bernstein EL, Jetton AE, Matsumoto Si, Markuns JF, Lehman MN, Bittman EL. Effects of Suprachiasmatic Transplants on Circadian Rhythms of Neuroendocrine Function in Golden Hamsters**This work was supported by NIH Grants MH-44132, KO2-MH-00914, and F32-HD-07673. A preliminary report of this research was presented at the 23rd Annual Meeting of the Society for Neuroscience (Neurosci Abstr 19:236.17, 1993). Endocrinology. 1999 Jan; 140(1):207–218. https://doi.org/10.1210/endo.140.1.6428, doi: 10.1210/endo.140.1.6428.

Michael AK, Fribourgh JL, Chelliah Y, Sandate CR, Hura GL, Schneidman-Duhovny D, Tripathi SM, Takahashi JS, Partch CL. Formation of a repressive complex in the mammalian circadian clock is mediated by the secondary pocket of CRY1. Proceedings of the National Academy of Sciences. 2017 Feb; 114(7):1560–1565. https://www.pnas.org/doi/full/10.1073/pnas.1615310114, doi: 10.1073/pnas.1615310114, number: 7 Publisher: Proceedings of the National Academy of Sciences.

Nakajima M, Imai K, Ito H, Nishiwaki T, Murayama Y, Iwasaki H, Oyama T, Kondo T. Reconstitution of Circadian Oscillation of Cyanobacterial KaiC Phosphorylation in Vitro. Science. 2005 Apr; 308(5720):414–415. https://doi.org/10.1126/science.1108451, doi: 10.1126/science.1108451, publisher: American Association for the Advancement of Science.

Nangle SN, Rosensweig C, Koike N, Tei H, Takahashi JS, Green CB, Zheng N. Molecular assembly of the period-cryptochrome circadian transcriptional repressor complex. eLife. 2014 Aug; 3:e03674. https://doi.org/10.7554/eLife.03674, doi: 10.7554/eLife.03674, publisher: eLife Sciences Publications, Ltd.

Parlak GC, Camur BB, Gul S, Ozcan O, Baris I, Kavakli IH. The secondary pocket of cryptochrome 2 is important for the regulation of its stability and localization. Journal of Biological Chemistry. 2022 Sep; 298(9). https://doi.org/10.1016/j.jbc.2022.102334, doi: 10.1016/j.jbc.2022.102334, publisher: Elsevier.

Partch CL. Orchestration of Circadian Timing by Macromolecular Protein Assemblies. Circadian Regulation: from Molecules to Physiology. 2020 May; 432(12):3426–3448. https://www.sciencedirect.com/science/article/pii/S0022283620300322, doi: 10.1016/j.jmb.2019.12.046.

Pelham JF, Dunlap JC, Hurley JM. Intrinsic disorder is an essential charac-

teristic of components in the conserved circadian circuit. Cell Communication and Signaling. 2020 Nov; 18(1):181. https://doi.org/10.1186/s12964-020-00658-y, doi: 10.1186/s12964-020-00658-y, number: 1.

Preitner N, Damiola F, Lopez-Molina L, Zakany J, Duboule D, Albrecht U, Schibler U. The Orphan Nuclear Receptor REV-ERB Controls Circadian Transcription within the Positive Limb of the Mammalian Circadian Oscillator. Cell. 2002 Jul; 110(2):251–260. https://doi.org/10.1016/S0092-8674(02)00825-5, doi: 10.1016/S0092-8674(02)00825-5, publisher: Elsevier.

Sadaie W, Harada Y, Matsuda M, Aoki K. Quantitative In Vivo Fluorescence Cross-Correlation Analyses Highlight the Importance of Competitive Effects in the Regulation of Protein-Protein Interactions. Molecular and Cellular Biology. 2014; 34(17):3272–3290. https://mcb.asm.org/content/34/17/3272, doi: 10.1128/MCB.00087-14, publisher: American Society for Microbiology Journals _eprint: https://mcb.asm.org/content/34/17/3272.full.pdf.

Scheiermann C, Gibbs J, Ince L, Loudon A. Clocking in to immunity. Nature Reviews Immunology. 2018 Jul; 18(7):423–437. https://doi.org/10.1038/s41577-018-0008-4, doi: 10.1038/s41577-018-0008-4.

Schmalen I, Reischl S, Wallach T, Klemz R, Grudziecki A, Prabu J Benda C, Kramer A, Wolf E. Interaction of Circadian Clock Proteins CRY1 and PER2 Is Modulated by Zinc Binding and Disulfide Bond Formation. Cell. 2014 May; 157(5):1203–1215. https://linkinghub.elsevier.com/retrieve/pii/S0092867414005352, doi: 10.1016/j.cell.2014.03.057, number: 5.

Schönle A, Von Middendorff C, Ringemann C, Hell SW, Eggeling C. Monitoring triplet state dynamics with fluorescence correlation spectroscopy: Bias and correction. Microscopy Research and Technique. 2014; 77(7):528–536. https://onlinelibrary.wiley.com/doi/abs/10.1002/jemt.22368, doi: 10.1002/jemt.22368, number: 7 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jemt.22368.

Seabold S, Perktold J. statsmodels: Econometric and statistical modeling with python. In: *9th Python in Science Conference*; 2010. .

Smyllie NJ, Bagnall J, Koch AA, Niranjan D, Polidarova L, Chesham JE, Chin JW, Partch CL, Loudon ASI, Hastings MH. Cryptochrome proteins regulate the circadian intracellular behavior and localization of PER2 in mouse suprachiasmatic nucleus neurons. Proceedings of the National Academy of Sciences. 2022 Jan; 119(4):e2113845119. https://doi.org/10.1073/pnas.2113845119, doi: 10.1073/pnas.2113845119, publisher: Proceedings of the National Academy of Sciences.

Vielhaber E, Eide E, Ann R, Zhong-Hua G, Virshup DM. Nuclear Entry of the Circadian Regulator mPER1 Is Controlled by Mammalian Casein Kinase I . Molecular and Cellular Biology. 2000 Jul; 20(13):4888–4899. https://doi.org/10.1128/MCB.20.13.4888-4899.2000, doi: 10.1128/MCB.20.13.4888-4899.2000, publisher: American Society for Microbiology.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020; 17:261–272. doi: 10.1038/s41592-019-0686-2.

Vitaterna MH, King DP, Chang AM, Kornhauser JM, Lowrey PL, McDonald JD, Dove WF, Pinto LH, Turek FW, Takahashi JS. Mutagenesis and Mapping of a Mouse Gene, Clock, Essential for Circadian Behavior. Science. 1994 Apr; 264(5159):719–725. https://doi.org/10.1126/science.8171325, doi: 10.1126/science.8171325, publisher: American Association for the Advancement of Science.

Wallach T, Schellenberg K, Maier B, Kalathur RKR, Porras P, Wanker EE, Futschik ME, Kramer A. Dynamic Circadian Protein–Protein Interaction Networks Predict Temporal Organization of Cellular Functions. PLOS Genetics. 2013 Mar; 9(3):e1003398. https://doi.org/10.1371/journal.pgen.1003398, doi: 10.1371/journal.pgen.1003398, publisher: Public Library of Science.

Wong DC, O'Neill JS. Non-transcriptional processes in circadian rhythm generation. Circadian Rhythms. 2018 Oct; 5:117–132. https://www.sciencedirect.com/science/article/pii/S2468867318301342, doi: 10.1016/j.cophys.2018.10.003.

Xing W, Busino L, Hinds TR, Marionni ST, Saifee NH, Bush MF, Pagano M, Zheng N. SCFFBXL3 ubiquitin ligase targets cryptochromes at their cofactor pocket. Nature. 2013 Apr; 496(7443):64–68. https://doi.org/10.1038/nature11964, doi: 10.1038/nature11964.

Xu H, Gustafson CL, Sammons PJ, Khan SK, Parsley NC, Ramanathan C, Lee HW, Liu AC, Partch CL. Cryptochrome 1 regulates the circadian clock through dynamic interactions with the BMAL1 C terminus. Nature Structural & Molecular Biology. 2015 Jun; 22(6):476–484. https://doi.org/10.1038/nsmb.3018, doi: 10.1038/nsmb.3018.

Yagita K, Yamaguchi S, Tamanini F, van der Horst GTJ, Hoeijmakers JHJ, Yasui A, Loros JJ, Dunlap JC, Okamura H. Dimerization and nuclear entry of mPER proteins in mammalian cells. Genes & Development. 2000 Jun; 14(11):1353–1363. http://genesdev.cshlp.org/content/14/11/1353.abstract, doi: 10.1101/gad.14.11.1353.

Yang N, Smyllie NJ, Morris H, Gonçalves CF, Dudek M, Pathiranage DRJ, Chesham JE, Adamson A, Spiller DG, Zindy E, Bagnall J, Humphreys N, Hoyland J, Loudon ASI, Hastings MH, Meng QJ. Quantitative live imaging of Venus::BMAL1 in a mouse model reveals complex dynamics of the master circadian clock regulator. PLOS Genetics. 2020 Apr; 16(4):e1008729. https://dx.plos.org/10.1371/journal.pgen.1008729, doi: 10.1371/journal.pgen.1008729, number: 4.

Yoo SH, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, Buhr ED, Siepka SM, Hong HK, Oh WJ, Yoo OJ, Menaker M, Takahashi JS. PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc Natl Acad Sci U S A. 2004; 101(15):5339–46. https://www.ncbi.nlm.nih.gov/pubmed/14963227, doi: 10.1073/pnas.0308709101, type: Journal Article.

Yoo SH, Kojima S, Shimomura K, Koike N, Buhr ED, Furukawa T, Ko CH, Gloston G, Ayoub C, Nohara K, Reyes BA, Tsuchiya Y, Yoo OJ, Yagita K, Lee C, Chen Z, Yamazaki S, Green CB, Takahashi JS. Period2 3-UTR and microRNA-24 regulate circadian rhythms by repressing PERIOD2 protein accumulation. Proceedings of the National Academy of Sciences. 2017 Oct; 114(42):E8855–E8864. https://doi.org/10.1073/pnas.1706611114, doi: 10.1073/pnas.1706611114, publisher: Proceedings of the National Academy of Sciences.

Yuan Q, Metterville D, Briscoe AD, Reppert SM. Insect Cryptochromes: Gene Duplication and Loss Define Diverse Ways to Construct Insect Circadian Clocks. Molecular Biology and Evolution. 2007 Apr; 24(4):948–955. https://doi.org/10.1093/molbev/msm011, doi: 10.1093/molbev/msm011.

Zheng B, Albrecht U, Kaasik K, Sage M, Lu W, Vaishnav S, Li Q, Sun ZS, Eichele G, Bradley A, Lee CC. Nonredundant Roles of the mPer1 and mPer2 Genes in the Mammalian Circadian Clock. Cell. 2001 Jun; 105(5):683–694. https://doi.org/10.1016/S0092-8674(01)00380-4, doi: 10.1016/S0092-8674(01)00380-4, publisher: Elsevier.

Zheng B, Larkin DW, Albrecht U, Sun ZS, Sage M, Eichele G, Lee CC, Bradley A. The mPer2 gene encodes a functional component of the mammalian circadian clock. Nature. 1999 Jul; 400(6740):169–173. https://doi.org/10.1038/22118, doi: 10.1038/22118.

Zhou HX, Bates PA. Modeling protein association mechanisms and kinetics. Catalysis and regulation / Protein-protein interactions. 2013 Dec; 23(6):887–893. https://www.sciencedirect.com/science/article/pii/S0959440X13001176, doi: 10.1016/j.sbi.2013.06.014.

Supplementary materials

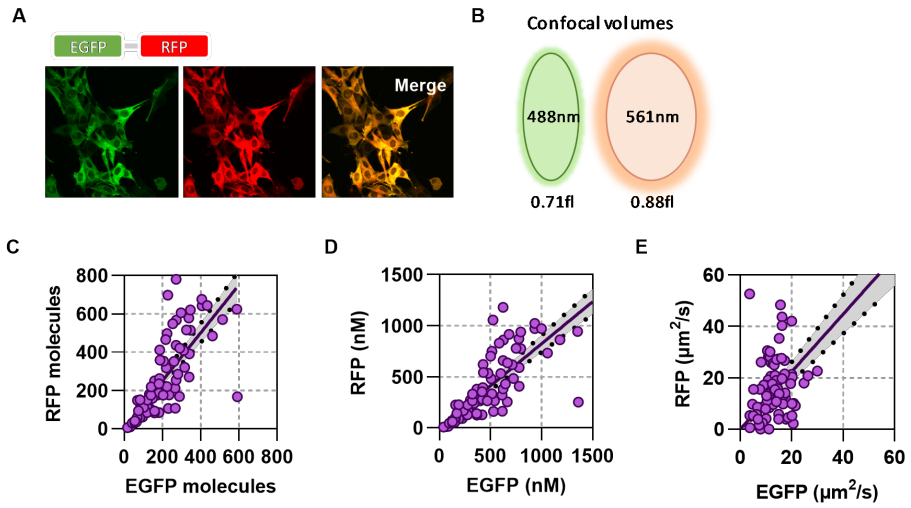

Figure supplement 1: Calibration of FCS using a direct fusion of two fluorescent proteins. (**A**) MEF cells were transduced to express EGFP and tagRFP as a single fused protein. FCS measurements of EGFP::tagRFP fusion were performed on MEF cells. Data was fit using a correlation model incorporating normal diffusion and an offset. (**B**) Illustration showing how confocal volume differs depending on the wavelength of the excitation laser, affecting calculation of number of molecules and diffusion across colours. Green EGFP and red tagRFP fluorescence is excited by 488 nm and 561 nm laser light respectively. Plots show (**C**) the number of molecules measured, (**D**) concentration calculated using the theoretical confocal volumes for each colour channel, and (**E**) diffusion of each protein.

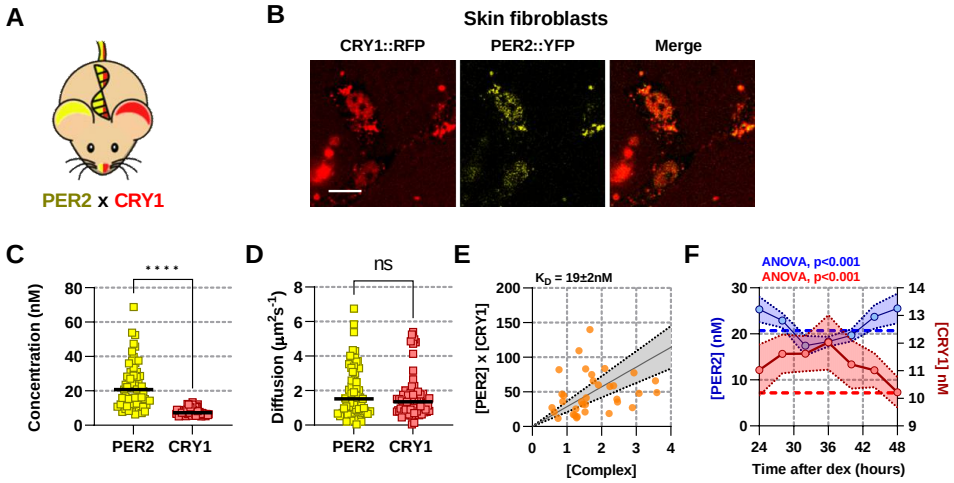

Figure supplement 2: Analysis of isolated skin fibroblasts from PER2 x CRY1 fluorescent knock-in mice without synchronising stimuli. (**A**) Knock-in mice were culled and isolated cells grown for two weeks prior to (**B**) confocal imaging. Image shows two cells expressing CRY1::mRuby3 and PER2::Venus nuclear fluorescence in separate channels as well as a merge channel. FC(C)S of nuclear fluorescence of PER2::Venus and CRY1::mRuby3 quantified (**C**) concentration, (**D**) diffusion rates of both proteins, and (**E**) disassociation constant K_D as a measure of interaction affinity. (**F**) Concentration data for unstimulated cells is plotted as a dotted line compared to seven time-point measurements made in dexamethasone synchronised cells replotted from Figure 1F. Statistical analysis was carried out using Kruskal-Wallis (values are denoted as $p > 0.05$ ns and $p < 0.0001$ ****).

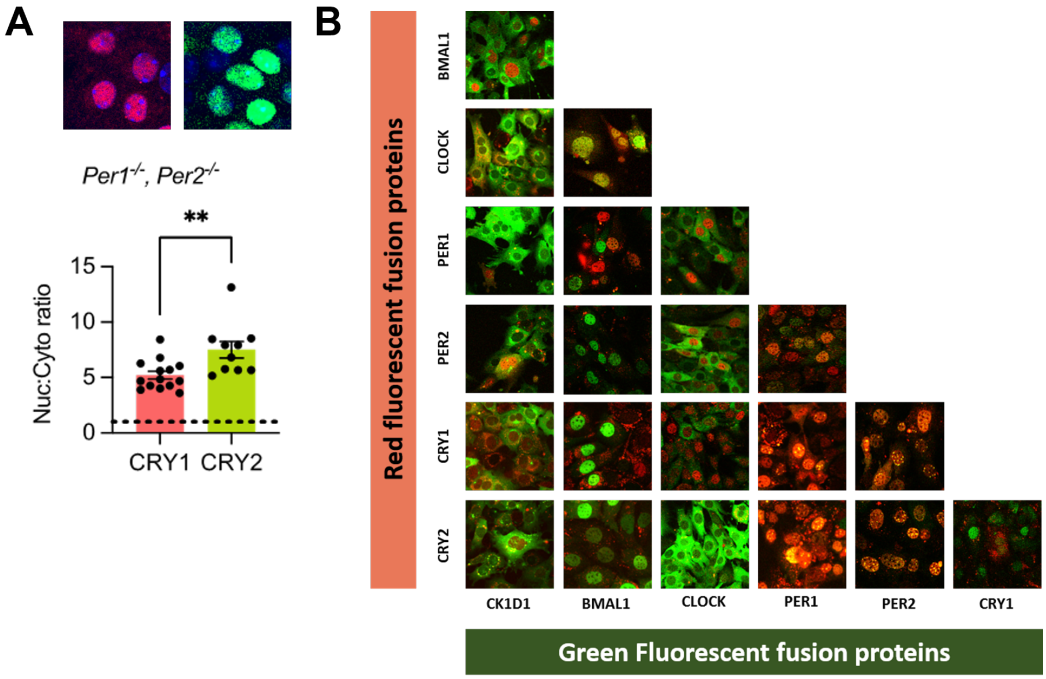

Figure supplement 3: (**A**) Localisation of CRY1::mRuby3 and CRY2::EGFP in single cells from organotypic SCN slices derived from PER1-/- PER2-/- global KO mice. CRY1::mRuby3 and CRY2::EGFP delivered into cells via transduced with adenovirus. (**B**) Merge confocal microscopy images of mouse embryonic fibroblasts expressing lentivirus transduced pairs of circadian proteins fused with either green EGFP or red tagRFP fluorescent proteins.

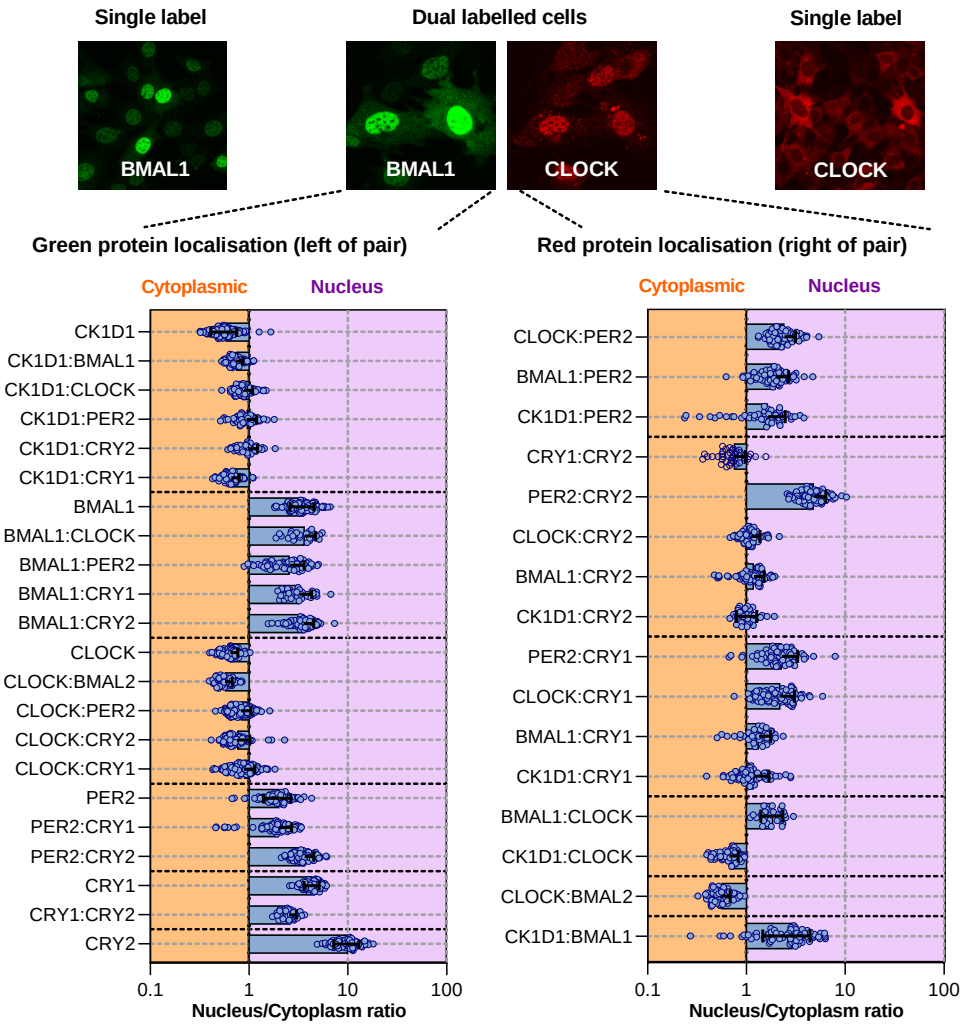

Figure supplement 4: Nuclear/cytoplasic ratios in MEFs Nuclear/Cytoplasmic ratio data from all available cell lines within this study. Top panel shows representative confocal microscopy images of cells transduced to express a single green fluorescent protein or red fluorescent protein as well as cells expressing both simultaneously. Bar graphs quantify nuclear/cytoplasmic ratio of fluorescence calculated for cells co-expressing two fluorescent proteins, showing the green fluorescent protein ratios on the left and red fluorescent protein ratios on the right. Values above 1 are more nuclear than cytoplasmic.

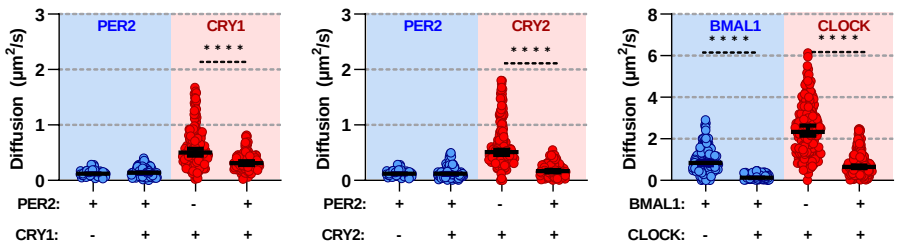

Figure supplement 5: Co-expression reduces the mobility of binding partners FCCS measurement of protein mobility performed in MEF cells transduced to express fluorescent PER2 and CRY paralogs alone or together as well as BMAL1 and CLOCK alone or together (denoted with '+' for transduced). Statistical analysis was carried out using Kruskal-Wallis (values are denoted as $p < 0.001$ *** and $p < 0.0001$ ****).

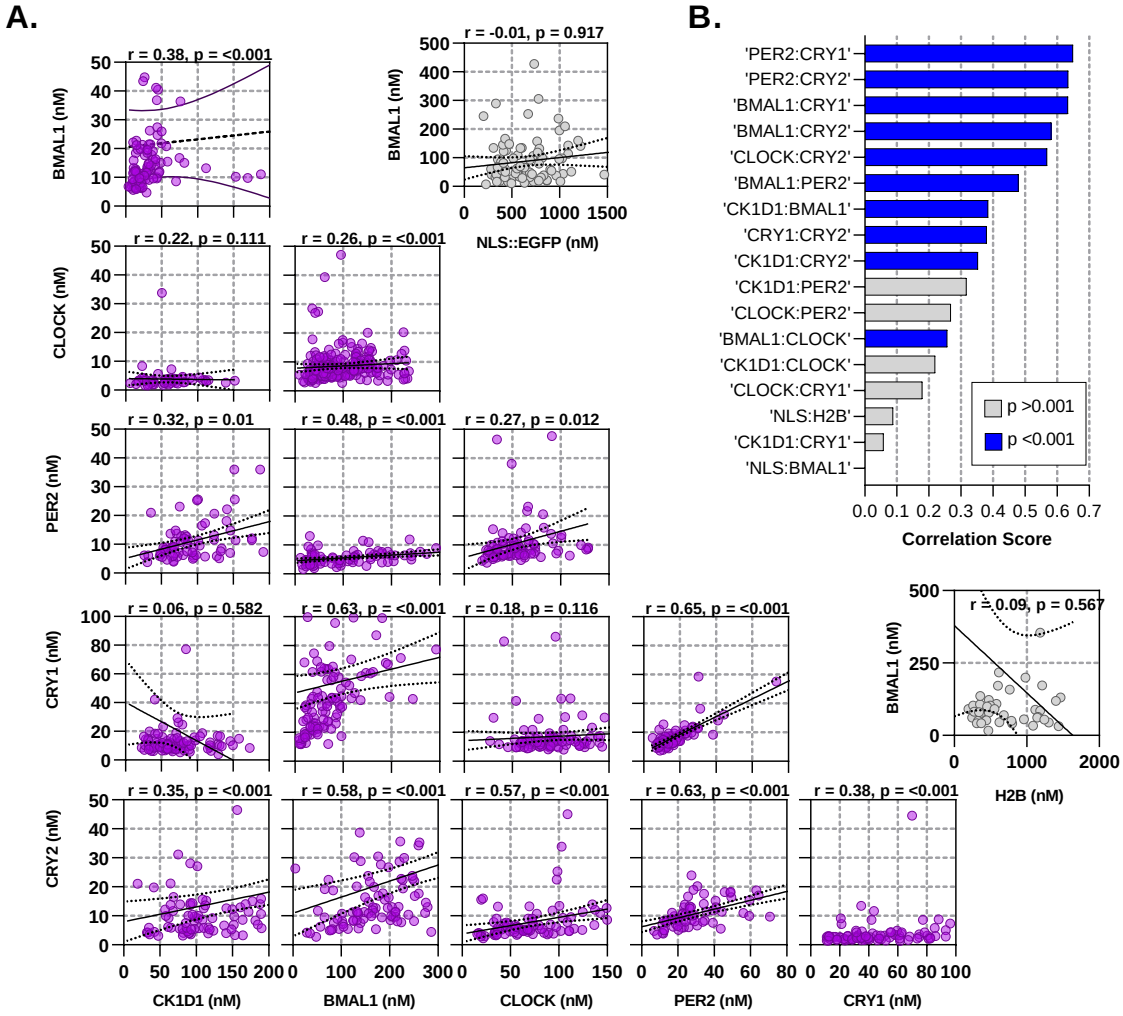

Figure supplement 6: Protein concentration correlations in lenti-viral MEFs Correlations between nuclear protein in MEF cells sequentially transduced to over-express pairs of fluorescent fusion proteins. (**A**) FCCS was performed on the nucleus of each cell line to determine concentration of protein showing concentrations of the EGFP and tagRFP fusions on horizontal and vertical axes, respectively. Each scatter plot shows linear regression model fits. (**B**) Spearman correlation was applied to all data and correlation score and p-values calculated.

Chapter 4

Improving analysis of fluorescence correlation spectroscopy

4.1 Overview

Molecular movement is inherently stochastic and as such is naturally described using probability theory. The probability behind the individual processes that describe a fluorescent molecule diffusing from one place to another and then emitting a photon that arrives at the detector is known. The difficulty here is to combine this knowledge into a complete probabilistic model of an FCS experiment connecting the observed photon arrival times to the parameters governing diffusion and subsequent photon emission. Jazani et al. (2019), Tavakoli et al. (2020), and colleagues in the Pressé group at Arizona State University proposed an alternative framework for analysing FCS data using Bayesian non-parametrics to define a likelihood that could then be found from observed photon data using Markov Chain Monte Carlo (MCMC). This allowed them to accurately infer molecular number, diffusion, and emission parameters directly from photon counts given enough computation time. However, these methods are difficult to use in practice on bulk data due to the $> 10^6$ MCMC iterations required to achieve convergence on most FCS datasets, a task that often takes many hours on a desktop computer. Their algorithm is computationally slow as it directly simulates approximately 50 molecules over plausible molecular trajectories over 1000s of time steps for each of the MCMC iterations prior to selecting which trajectories contribute to the likelihood.

Here I lay out the probabilistic underpinnings of a model of an FCS experiment and derive how the variance of changes in photon counts is dependent on the molecular parameters. Through Monte Carlo simulations of FCS experiments (see section 4.2 for details) and theoretical arguments, I find in section 4.4 that distributions of intermediate variables linking parameters to data to be well-behaved and readily approximated. This motivated the work in section 4.2 where I derive an approximate likelihood of observing the photon data parameterised by the diffusion rate, molecular number, and per molecule emission rate. My maximum approximate likelihood method (FCS-MAL) replaces the slow computation step found in the algorithms of Jazani et al. (2019) and Tavakoli et al. (2020) with an equation that may be maximised to infer the molecular parameters orders of magnitude faster with minimal loss of accuracy.

4.2 Journal paper: Accelerated analysis of fluorescence correlation spectroscopy

Title	A parametric probabilistic model for Fluorescence Correlation Spectroscopy
Journal	Manuscript in preparation
Year	N/A
Authors	**Alex A Koch**, James S Bagnall, Bailey Winstanley, David G Spiller Korbinian Strimmer, Andrew SI Loudon, Thomas A House
doi	N/A

4.2.1 My contributions

I conceived of and developed this work with James Bagnall, David Spiller, Korbinian Strimmer and Thomas House. Mathematical calculations were completed by me with help from Bailey Winstanley, Korbinian Strimmer, and Thomas House. Along with James Bagnall I prepared fluorescent cells and performed FCS measurements. Finally, all computational work was carried out by myself.

Article

FCS-MAL: Accelerated analysis of fluorescence correlation spectroscopy

Alex A Koch[1], James S Bagnall[1], Bailey Winstanley[2], David G Spiller[1], Korbinian Strimmer[3], Andrew S I Loudon[1], and Thomas A House[3,*]

[1]Faculty of Biology, Medicine and Health, University Manchester, UK
[2]Department of Physics and Astronomy, University of Manchester, UK
[3]Department of Mathematics, University of Manchester, UK
[*]Correspondence: thomas.house@manchester.ac.uk

ABSTRACT Fluorescence Correlation Spectroscopy (FCS) is a widely employed microscopy technique to quantitatively derive concentration and diffusion coefficients from fluorescent molecules diffusing within a small volume illuminated by focused light. Standard FCS requires relatively high illumination and/or measurement times, exacerbating photobleaching issues, limiting its use in the life sciences. Kinetic parameters of these fluorescent molecules are inferred from time-lagged correlations which effectively averages large amounts available data, reducing data density to as low as $\sim 10^2$ correlation points from the original $\sim 10^5$ photon arrival times. Furthermore, commonly applied least-squares curve fitting approaches to these data are prone to poor fits and underestimation of uncertainty on fitted parameters. Here, we derive a probabilistic likelihood model for a typical FCS experiment from first principles, and demonstrate how dynamical parameters may be inferred directly from information rich photon arrival time data by maximising this approximate likelihood. We denote this FCS-Maximum approximate likelihood (FCS-MAL). This approach allows simultaneous and rapid evaluation of concentration, diffusion coefficient, and molecular brightness from fewer data points than other FCS analyses.

SIGNIFICANCE Fluorescence correlation spectroscopy analysis is limited to high signal to noise and relatively long measurement times. Here we demonstrate how diffusion rates, molecular numbers, and molecular brightness of fluorescent molecules may be quantified directly from unprocessed single photon arrival times collected during an FCS experiment to improve inferences whilst reducing required measurement times. Our approximate likelihood of observing these photons is both as accurate and computationally quicker than both FCS and new Bayesian non-parameteric approaches.

Introduction

Fluorescence Correlation Spectroscopy (FCS) is a widely utilised microscopy technique that quantifies the concentration, diffusion, and interactions of fluorescent molecules. With its high temporal resolution, ranging from nanoseconds to seconds, FCS sensitively records fluctuations in photon intensity resulting from single molecules entering and exiting a small confocal volume of approximately 1 femtoliter, Figure 1A. The first application of FCS was in the analysis of fluorescent dyes binding to DNA in solution (Magde et al., 1972), and since then it has become a crucial tool in the life sciences for the study of fluorescently labelled molecules and proteins in vitro and in vivo (Schwille, 2001; Yu et al., 2021).

FCS works by using the non-linearities in the optics of the microscope to gather time-resolved information about the fluorescent molecules. The confocal microscope is setup so that the illumination is unevenly distributed, with the greatest likelihood of detecting a photon located at center of the illuminated area (Qian, 1990), Figure 1B. Fluorescent molecules that enter this volume become brighter as they move towards the center of the beam and fade as they leave, providing information about their behaviour (Elson and Magde, 1974), Figure 1C. The emitted photons contain dynamic information about the fluorescent molecules, which is used to estimate important properties, such as concentration, diffusion (Höfling et al., 2011), and protein-protein binding affinity (Bacia et al., 2006; Sadaie et al., 2014). This information is extracted by auto-correlating the fluctuating intensity traces over different lag-times and then fitting theoretical models via non-linear regression, Figure 1D,E. These models have been used to study a variety of molecular properties, including normal and anomalous diffusion coefficients (Höfling et al., 2011), chemical relaxation times (Haupts et al., 1998), photochemical triplet states (Widengren et al., 1995), and protein-DNA binding kinetics (Michelman-Ribeiro et al., 2009).

Traditional FCS analyses face a number of technical challenges that arise from the experimental setup and analysis. These challenges are primarily due to the use of multi-tau cor-

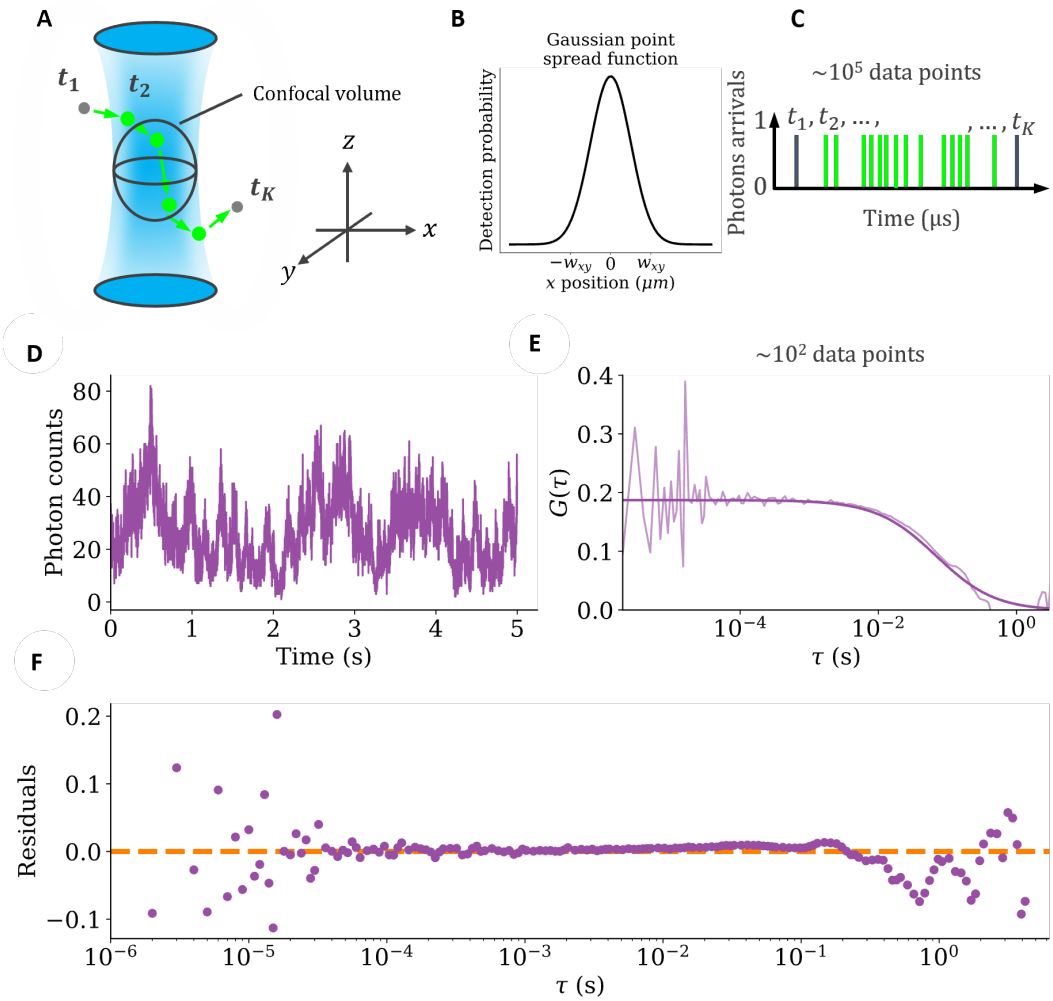

Figure 1: Fluorescence correlation spectroscopy resolves single photon arrivals. (**A**) Schematic of an FCS experiment. Fluorescent molecules (grey and green) diffuse in and out of the laser illuminated (blue) confocal volume fluorescing (green) with increasing strength closer to the centre over time steps sampled from $k = 1$ to $k = K$. (**B**) One dimensional slice of a Gaussian point spread function that defines the probability over space of emission and detection of a photon. This is mirrored for the y and z dimensions. (**C**) Photons arriving at the detector (green) are resolved through ultra-fast 10^6 Hz sampling at the detector, carrying with them dynamical information about the particles which emitted them. (**D**) Millisecond-binned time trace of photon detection for a slow diffusing ($0.1\mu m^2 s^{-1}$) molecule at a concentration of (17.30 nM). (**E**) Photon emissions are well correlated over short time intervals, becoming increasingly less correlated over longer times as particles diffuse out of the volume. FCS analysis fits theoretical correlation models (dark line) through these data to obtain estimates for the dynamical parameters of the system. (**F**) Residuals for the fit to data in (**E**) demonstrates increased uncertainty (noise) at short ($\tau < 10^{-4}$ s) and long ($\tau > 10^{-1}$ s) lag times.

relators, intrinsic noise, and non-linear regression. In the early stages of FCS development, calculating the auto-correlation of photon intensities was a computationally intensive task, leading to the popularity of multi-tau correlators as a solution. These correlators reduced the computational demands, making FCS analysis more efficient and accessible (Schwille et al., 1999; Rigler et al., 1993). Multi-tau correlators work by sub-sampling the auto-correlation by averaging over increasingly wider time bins of logarithmic density (Schätzel et al., 1988). This compresses the data, leading to loss of information and introduction of systematic uncertainty, (Saffarian and Elson, 2003). Multi-tau correlation vastly reduces the number of data points available for fitting and parameter inference from around $10^5 - 10^6$ photon arrivals to $\sim 10^2$ auto-correlation values. These correlators can also exacerbate the large variance in the auto-correlations at short and long lag-times, see Figure 1F. At very short lag-times the density of data is low due to sampling rates being comparable with average emission rates (shot noise). Whereas at long lag-times the 'particle noise' becomes dominant and results from the observed photons becoming less likely to originate from the same particles (Koppel, 1974; Qian, 1990; Kask

et al., 1997; Saffarian and Elson, 2003). Finally, the challenges posed by high variance and a limited number of data points are exacerbated by the prevalent use of non-linear least squares regression in fitting these correlation curves. While regularisation schemes can be incorporated during fitting to counter problems, such as bias towards outliers (Koch et al., 2022), least squares regression does not provide insights into alternative solutions, if they exist (Stuart, 2010). As a result, traditional FCS analyses reliant on least squares regression might offer constrained accuracy and reliability. A promising solution to these hurdles in traditional FCS analyses is the introduction of a new analytical method that draws on Bayesian or maximum likelihood techniques and leverages the raw photon arrival data.

Despite its first introduction by Magde et al. (1972), FCS remains technically challenging to implement in biological applications due to several limitations. To obtain sufficient data without bleaching or causing photo-toxic damage to cells (Magidson and Khodjakov, 2013), measurements must be taken over relatively long periods of time (> 1s), which limits the technique to bright or high-concentration molecules. The low abundance of many proteins, such as transcription factors, results in a low photon emission rate when fluorescently labelled, exacerbating systematic errors as the signal becomes comparable to noise. Given these technical and biological limitations, there is a need for enhanced FCS analysis techniques that are sensitive, accurate, reliable, and require less acquisition time.

To address many of the technical challenges faced in FCS, Jazani et al. (2019) and Tavakoli et al. (2020) from the Pressé group proposed an alternative framework based on Bayesian non-parametrics to quantify the distribution of the molecular number and diffusion rates from FCS data. As their model relating distributions over molecular positions and collected photons cannot be integrated analytically, they used Markov Chain Monte Carlo (MCMC) to compute this likelihood integral by simulating multiple plausible molecular trajectories and selecting the trajectories that best explain the data. The diffusion rates and concentrations are then inferred from these trajectories. This approach is advantages of accurately inferring low numbers of molecules from short time traces (~ 1 ms) and rigorous quantification of uncertainties. However, when analysing multiple FCS experiments on a typical computer, this method can be prohibitively slow.

We aimed to address the limitations and challenges in FCS by calculating a likelihood function that was parameterised by molecular number and diffusion coefficient. To do so, we applied carefully controlled approximations, including Laplace's method (Murray, 1984), to compute the likelihood function in closed form. We then maximised this approximate likelihood equation for the full unprocessed photon arrival time data, resulting in a new method we denote as maximum approximate likelihood (MAL). Our FCS-MAL approach takes advantage of the increased data density to extract molecular parameters from FCS experiments with improved accuracy and reliability, using simple formulae. Our proposed method yields the same high-accuracy inferences from short measurements as the alternative Bayesian non-parametric framework proposed by the Pressé group (Jazani et al., 2019; Tavakoli et al., 2020), while still maintaining the fast computation time of traditional FCS analysis.

Materials and Methods

Model formulation

Our goal is to estimate the parameters defining an FCS experiment, primarily the diffusion coefficient, D, and number of molecules in the confocal volume, N. Mirroring the approach of Tavakoli et al. (2020), we wish to know the posterior distribution, $P(\theta|\boldsymbol{\Delta t})$, of the parameters θ given a data set of times between detected photons $\boldsymbol{\Delta t} = (\Delta t_1, \Delta t_2, \ldots, \Delta t_{K-1})$ for $\Delta t_k = t_{k+1} - t_k$. To compute the posterior, we require the likelihood, $P(\boldsymbol{\Delta t}|\theta)$, of observing the data given the parameters, which is impossible to solve in closed form. Hence, our aim is to approximate the likelihood in terms of the parameters θ. For normal Brownian motion in the absence of background photon emission, θ is the set $\{\mu_{\mathrm{m}}, D, N, w_{xy}, w_z\}$ where w_{xy} and w_z define the size of the confocal volume, and the per-molecule emission rate is given by μ_{m}. The number of detected photons depends on the positions of the molecules and their diffusion rates. Diffusion defines how the number of photons collected at the detector change over time as molecules move in and out of the confocal volume. Thus, to construct our model we need to expand the likelihood in terms of the molecular positions $\underline{\underline{\boldsymbol{r}}}$. Here, a single underline $\underline{\boldsymbol{r}}_k$ defines the set of 3D positions, $\boldsymbol{r}_{n,k} = (x_{n,k}, y_{n,k}, z_{n,k})$, at time step k for N total molecules, such that a double underline, $\underline{\underline{\boldsymbol{r}}}$, defines the full set of positions for N molecules over K time steps. It can be shown (see appendix for full details of calculations) that the likelihood is a marginal integral over all positions

$$P(\boldsymbol{\Delta t}|\theta) = \int_{\underline{\underline{\boldsymbol{r}}}} P(\boldsymbol{\Delta t}|\underline{\underline{\boldsymbol{r}}}, \theta) P(\underline{\underline{\boldsymbol{r}}}|\theta) d\underline{\underline{\boldsymbol{r}}}. \tag{1}$$

By assuming that photons at time k only depend on the positions and parameters at time k, the first term becomes a product of the probability of observing each time interval between photons, Δt_k,

$$P(\boldsymbol{\Delta t}|\underline{\underline{\boldsymbol{r}}}, \theta) = \prod_{k=1}^{K-1} P(\Delta t_k|\underline{\boldsymbol{r}}_k, \theta). \tag{2}$$

As normal diffusion is a Markovian process, the second term can be similarly expanded over each time step

$$P(\underline{\underline{\boldsymbol{r}}}|\theta) = \left(\prod_{k=1}^{K-1} P(\underline{\boldsymbol{r}}_{k+1}|\underline{\boldsymbol{r}}_k, \theta)\right) P(\underline{\boldsymbol{r}}_1|\theta), \tag{3}$$

for an initial set of molecular positions $\underline{\boldsymbol{r}}_1$. In this study, we assume that $\underline{\boldsymbol{r}}_1$ is uniformly distributed and we consider molecules diffusing through normal Brownian motion, such that the positions in each direction update according to a normal distribution. Thus $P(\underline{\boldsymbol{r}}_{k+1}|\underline{\boldsymbol{r}}_k, \theta)$ is a multivariate normal distribution with mean centered on the previous position $\underline{\boldsymbol{r}}_k$, and variance dependent on diffusion rate D as $2D\Delta t_k$. In 3D and without complex fluorophore dynamics such as blinking caused by triplet states, $P(\Delta t_k|\underline{\boldsymbol{r}}_k, \theta)$ follows an exponential distribution with parameter

$$\mu_k = \mu_m \sum_{n=1}^{N} \mathrm{PSF}(\boldsymbol{r}_{n,k}). \tag{4}$$

The point spread function (PSF) determines the normalised probability of a molecule fluorescing under focused laser light and is defined by the optics of the system (Schwille, 2001). The molecular brightness μ_m is the rate of photons emission per second and is a function of the excitation probability, quantum yield, detection efficiency, and laser intensity (Chen et al., 1999). In this study, we have chosen to complete the calculation of Equation 1 for optics defined by a 3D Gaussian point spread function (PSF) (Schwille, 2001; Zhang et al., 2007), Figure 1B,

$$\mathrm{PSF}_{3\mathrm{G}}(\boldsymbol{r}_{n,k}) = \exp\left[-2\left(\frac{x_{n,k}^2 + y_{n,k}^2}{w_{xy}^2} + \frac{z_{n,k}^2}{w_z^2}\right)\right]. \tag{5}$$

Other common point spread functions in 3D include Lorentzian and Gaussian-cylindrical functions. In all these cases, the PSF contains an exponential term leading to a double exponential in $P(\Delta t_k|\underline{\boldsymbol{r}}_k, \theta)$ in at least one direction, i.e. e^{e^x}. These double exponential terms prohibit the likelihood integral from being solved analytically. Fortunately, for an appropriately large μ_k, $P(\Delta t_k|\underline{\boldsymbol{r}}_k, \theta)$ is highly peaked, as such we may apply the Laplace approximation (Murray, 1984), which transforms $P(\Delta t_k|\underline{\boldsymbol{r}}_k, \theta)$ into a readily integrable Gaussian function, Figure 2. This approximation is valid when $\Delta t < \frac{1}{\mu_m}$, i.e. there is on average more than one molecule within the confocal volume significantly contributing to the signal. The application of the Laplace approximation yields

$$P(\Delta t_k|\underline{\boldsymbol{r}}_k, \theta) \approx \mu_m \left(\sum_{n=1}^{N} \left(e^N \prod_{n'=1}^{N} \mathrm{PSF}_{3\mathrm{G}}(\boldsymbol{r}_{n',k}) \right)^{-\mu_m \Delta t_k} \times \mathrm{PSF}_{3\mathrm{G}}(\boldsymbol{r}_{n,k}) \right). \tag{6}$$

By applying Equation 6 to equation Equation 2, the likelihood can now be integrated over all positions for K times and N molecules using standard Gaussian integrals. See the appendix for full details. The final result of the combination of integrals is

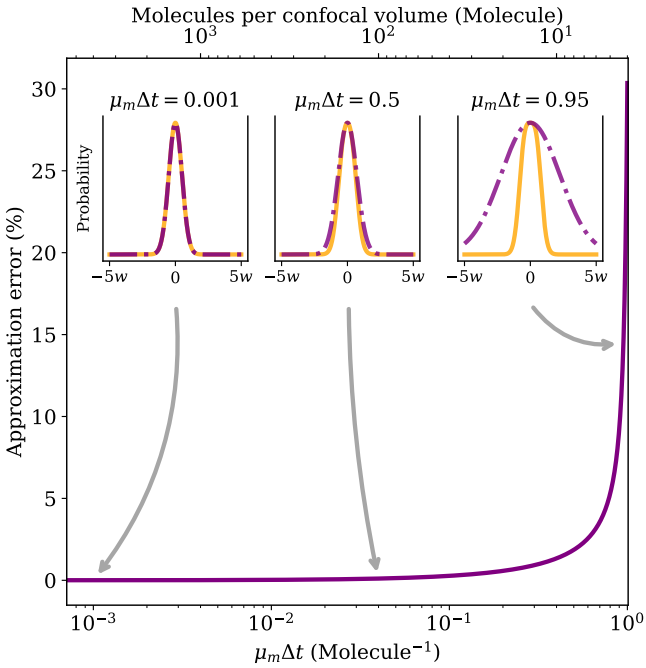

Figure 2: Laplace approximation is valid for frequent photon emissions. The Laplace approximation models the full double exponential function as a Gaussian function that is asymptotically equivalent to $P(\Delta t_k|\underline{\boldsymbol{r}}_k, \theta)$ as $\mu_m\Delta t \to 0$. Insets show the Laplace approximation in purple (dashed) and the original model in orange (solid) for a 1D Gaussian PSF for three values of $\mu_m\Delta t$ given a confocal width of $w = 0.2$ arb. units. For values of $\mu_m\Delta t < 0.3$ the percentage error between the approximation and full model over all space is below 1%. The number of molecules is given along the top axis and is proportional to $1/\mu_m\Delta t$.

$$P(\boldsymbol{\Delta t}|\theta) \approx \frac{\beta_{xy}^2 \beta_z}{V_{\mathrm{eff}}} \left(\prod_{k=1}^{K-1} \mu_m N^3 e^{-\mu_m N \Delta t_k} \right) \tag{7}$$

where the effective volume $V_{\mathrm{eff}} = \pi^{\frac{3}{2}} w_{xy}^2 w_z$ defines the size of the confocal volume within which observed photons are collected from and is dependent on the optical setup (see Experimental FCS setup for details). Outside of this volume the rate of photon emission from each particle is negligible and thus the number of molecules quantified by N is strictly the number of observed molecules within V_{eff}. This allows us to estimate the concentration of the fluorescent molecules in solution. Furthermore,

$$\beta_d = 1 - \frac{2D}{w_d^2} \sum_{k=1}^{K-1} \Delta t_k \left[N \left(\sum_{k'=1}^{k} \mu_m \Delta t_k \right) - k \right] - \frac{8D^2}{w_d^4} \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{K-1} ij \Delta t_i \Delta t_j + \sum_{k=1}^{K-1} (k-1)(\Delta t_k)^2 \right). \tag{8}$$

for $d = \{xy, z\}$. Note that the finite number of time indices k range from $k = 1$ to $k = K - 1$ due to the definition of $\Delta t_k = t_{k+1} - t_k$. This approximate likelihood may be used to estimate parameters of the model using Bayesian inference, alongside priors for each parameter in θ, or via maximum likelihood techniques. Here we demonstrate the efficacy of this likelihood model using maximum likelihood estimation.

Maximum approximate likelihood estimation

Estimators for each of the parameters $\hat{\theta}^{(i)}_{\text{MAL}}$ may be found by maximising the logarithm of the approximate likelihood in Equation 7. The estimators, $\hat{\theta}^{(i)}_{\text{MAL}}$, are defined such that they maximise the approximate log-likelihood. Hence, the maximum likelihood estimate can be found by solving for $\hat{\theta}^{(i)}_{\text{MAL}}$ when the derivative with respect to each estimator is zero, i.e. $\partial_{\theta_i} \ln(P(\boldsymbol{\Delta t}|\theta)) = 0$. We denote this procedure as Maximum Approximate Likelihood (MAL) estimation. MAL yields the estimator for Molar concentration,

$$\hat{C}_{\text{MAL}} = \frac{\hat{N}_{\text{MAL}}}{N_A V_{\text{eff}}} = \frac{3\varepsilon}{N_A V_{\text{eff}} \mu_{\text{m}} \langle \boldsymbol{\Delta t} \rangle}, \tag{9}$$

up to a correction factor ε, where N_A is Avrogadro's number. This also depends on the average difference in arrival times which is $\langle \boldsymbol{\Delta t} \rangle = \frac{1}{K-1} \sum_{k=1}^{K-1} \Delta t_k$. The factor, ε, corrects the effective volume V_{eff} and must be determined for each experiment. The estimator for the diffusion coefficient, D, is

$$\hat{D}_{\text{MAL}} = \frac{\alpha(K-1)}{4(w_{xy}^{-2} + 2w_z^{-2})\mu_m N \sum_{i=0}^{K-1} \text{Cov}(\boldsymbol{\Delta t}_k, \boldsymbol{\Delta t}_{k+i})} + \eta, \tag{10}$$

where α and η are constants to be determined for each set of measurements. The covariance between Δt_k values is calculated over shifts, i.e. we are comparing values close to one another and then over progressively larger time intervals. This closely mirrors the traditionally correlative FCS analyses. Here, MAL has a natural intuitive interpretation that the average number of molecules, given in Equation 9, depends on the ratio of the average frequency of detected photons to the per-molecule emission rate. Furthermore, the diffusion rate D relies on how photon emissions vary over time. Standard errors on each of these estimators is calculated via bootstrapping such that each estimator is sampled multiple times from the available data with replacement using the *bootstrap* function from SciPy (version 1.11.1) (Virtanen et al., 2020; Efron et al., 1994).

Synthetic data generation

To validate our model, synthetic data was generated by a Python 3 program (FCS sample generator; github.com/LoudonLab/FcsAnalysisPipeline) that was based on the MATLAB algorithm created by Jazani et al. (2019). We simulated N molecules diffusing in 3D space, with positions drawn from a normal distribution with variance $2D\delta t$. At each time-step, δt, the number of emitted photons were drawn from a Poisson distribution with a mean dependent on the molecular brightness, μ_m, and point spread function of the molecular positions relative to the centre of the confocal volume. Times between photon arrivals, Δt_k, were calculated from the interval between these arrivals. Unless otherwise stated we performed simulations with similar parameters as Jazani et al. (2019), specifically $w_{xy} = 0.3$ µm, $w_z = 1.5$ µm, $\mu_m = 22500$ s^{-1} photons per molecule, and $\delta t = 10^{-7}$ s. The confocal widths, w_{xy} and w_z, are slightly larger than those present in our experiments, but allow for faster generation of synthetic data without violating any of the key assumptions of our model. See the appendix for a full description of the simulation scheme. Auto-correlations were calculated using the multiple tau Python package (Müller, 2012).

FCS analysis

Experimental FCS data extracted from the Zeiss .fcs files was analysed using the same software pipeline as previously described in Koch et al. (2022). In the case of synthetic data, the same software pipeline was utilised to fit models to autocorrelation curves obtained from generated .sim files.

Experimental FCS setup

For imaging experiments, cells were plated into 35 mm glass bottomed imaging dishes (Greiner Bio-One) at least 6 hours prior to imaging. FCS measurements were performed in the nucleoplasm of each cell over an acquisition time of 10-20 seconds on a Zeiss LSM 880 confocal microscope with the ConfoCorr3 module running ZEN 2.1 SP3 FP2 or ZEN 2010b SP1 software following the same protocol as (Koch et al., 2022). An effective volume of 1 airy unit was used after calibrating the pinhole in the x-y plane for maximal signal intensity. When using the FLUAR 40x NA 1.3 oil immersion objective, this volume is approximately 0.72 fL for 488 nm laser excitation. This effective confocal volume was calculated usting the equation

$$V_{\text{eff}} = \pi^{\frac{3}{2}} w_{xy}^2 w_z = \pi^{\frac{3}{2}} \left(\frac{0.61\lambda}{\text{NA}}\right)^2 \left(\frac{2n\lambda}{\text{NA}^2}\right), \tag{11}$$

where w_{xy} and w_z are the beam widths in the x, y, and z axes respectively. NA is the numerical aperture (NA = 1.3 for our 40x objective), λ the wavelength of the laser, and n the refractive index of the immersion oil ($n = 1.515$ in all experiments). Laser power was reduced to minimize photobleaching whilst maintaining counts per molecule greater than 0.3 kHz.

Fluorescent cell generation and maintenance

Mouse embryonic fibroblasts (MEF) (Yoo et al., 2017) derived from previously reported Per2::Luc mouse (Yoo et al., 2004) were cultured in Gibco DMEM/F-12 (Cat. No. 11320033) with 10% fetal bovine serum (HyClone). These cells were tested for the absence of mycoplasma using MycoAlert mycoplasma detection kit (Cat. No. LT07-418). Cells were passaged every 3-4 days, maintaining cells until passage 30. Production of 3rd generation lentivirus was performed according to the same protocol described in (Bagnall et al., 2015). At low passage (6-12) lentivirus transduction was performed to express a fusion of enhanced green fluorescent protein (EGFP).

Plasmids

Lentivirus transfer plasmid encoding for EGFP (pLNT-EGFP) was generated via the gateway cloning system as previously described (Bagnall et al., 2015). Vectors were recombined with a target destination vector containing a fluorescent protein sequence to generate a terminal lentivirus vector, in which expression is regulated from the constitutive ubiquitin ligase C promoter (UbC).

Results

Overview

Our aim is to infer molecular concentrations and diffusion coefficients from data collected in a typical FCS experiment. In the previous section we set out how to estimate these quantities from the intervals between photon arrival times at the detector, $\boldsymbol{\Delta t}$, from our approximate likelihood model using maximum approximate likelihood (MAL). We first validate our method using synthetic data from Monte Carlo simulations of an FCS experiment. We then demonstrate this method upon real in vivo measurements of fluorescent proteins in cells.

Validation by simulated data

We validated our model against synthetic data by finding estimates of molecular concentrations and diffusion rates using our maximum approximate likelihood technique (MAL). As expected, the simple formula in Equation 9 estimates molecular concentrations well when molecular brightness, μ_m, is known (Figure 3A). Mirroring traditional FCS analysis, the estimator $\hat{C}_{\mathrm{MAL}}$ is insensitive to the diffusion rate of the molecules. However, to calculate the concentration, we must know the effective confocal volume V_{eff} and it's correction factor ε. Here, following calibration over all of the data the correction factor was determined to be $\varepsilon = 1.39$. In experiments this corrective factor is determined from performing FCS calibration measurements on solutions of fluorophores with known concentrations such as recombinant fluorescent proteins or fluorescent reference dye (Politi et al., 2018). As expected, estimations of the molecular concentration improve as the number of data points increases. Strikingly, $\hat{C}_{\mathrm{MAL}}$ converges towards the true value after only 10^4 photon arrivals, which in this experiment is equivalent to a collection time of just 100 ms (Figure 3B). Similarly diffusion rates can be inferred using the estimator $\hat{D}_{\mathrm{MAL}}$, (Figure 3C). Calculated values for the diffusion coefficient also rapidly converge as data density increases (Figure 3D). Due to the unstable nature of the sums and ratios in $\hat{D}_{\mathrm{MAL}}$, we find that our data sets require a minimum of 10^3 data points to avoid infinities caused by divisions by zero.

Application to biological data

To demonstrate FCS-MAL on real biological data we transduced mouse embryonic fibroblasts (MEF) with a lentivirus carrying enhanced green fluorescent protein (EGFP) driven by the constitutive ubiquitin ligase C promoter (UbC) (Figure 4A). FCS measurements over 10 seconds were performed in the nucleus of 35 cells. These data gave MAL estimates of the concentration that were well correlated with those from standard FCS auto-correlation fits (Figure 4B). Remarkably, within single cells the concentration can be found with fewer than 10,000 data points, which corresponds to less than 0.1 s of measurement time (Figure 4).

Discussion

High resolution fluorescent microscopy reveals the dynamics of single molecules from information encoded in emitted photons. Correlative methods for quantifying abundance and movement of fluorescent molecules have seen extensive use since their inception (Magde et al., 1972; Yu et al., 2021). However, to obtain robust estimates of molecular number and diffusion rates with reasonable uncertainties, analyses using these auto-correlations require long individual traces and large data-sets. This is exacerbated for in vivo measurements where intrinsic biological noise, propensity for cell movement, fluorophore bleaching, and avoidance of phototoxic exposure places limits on measurements. As such, measurement times between $5 - 60$ s with over 30 replicate measurements are typically required to properly characterise a fluorescent protein in vivo.

These problems have been addressed by the use of Bayesian non-parametrics on raw FCS data by Jazani et al. (2019) and Tavakoli et al. (2020). They used MCMC methods regularised by a Kalman filter to computationally integrate the likelihood function within their Bayesian model. Whilst this method is accurate and provides rigorous quantification of uncertainties upon inferred parameters, it requires a large amount of MCMC iterations to converge, often over 10^6,

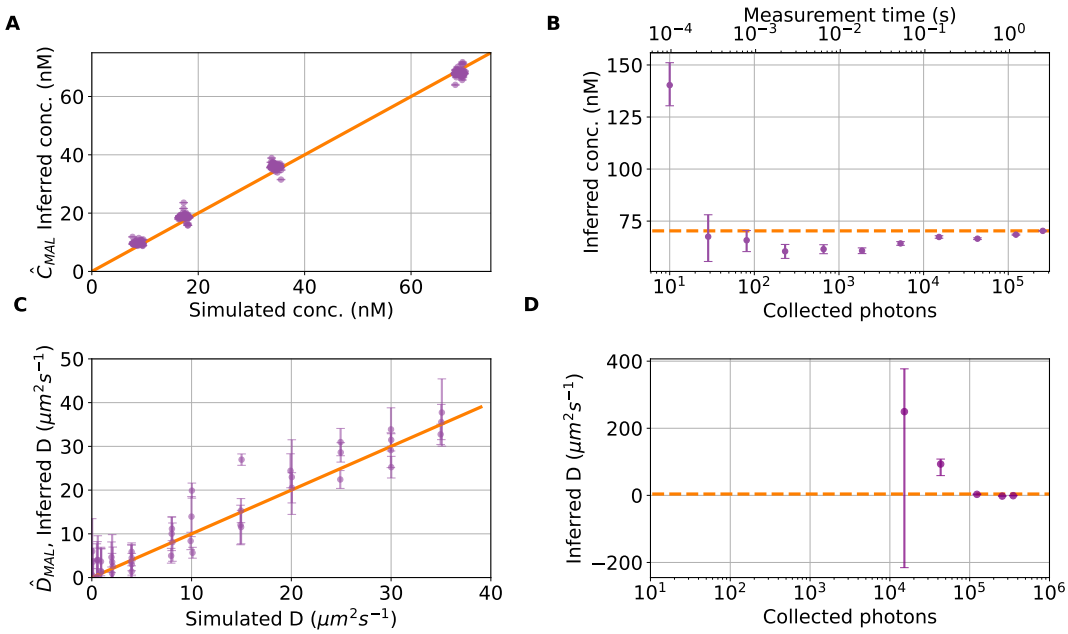

Figure 3: MAL accurately infers concentration and diffusion rates for synthetic data with only 10^4 collected photons. Inferred concentration of molecules from simulations over differing concentrations of molecules (8.65, 17.30, 34.60, 69.19 nM) and diffusion coefficients (0.1 – 35.0 $\mu m^2 s^{-1}$) for a fixed molecular brightness of $\mu_m = 22500$ (**A**) A one-to-one relationship of simulated to inferred concentration using the MAL estimator for $\hat{C}_{\mathrm{MAL}}$ found from Equation 9 over all available data. (**B**) Inferred concentration converges to true simulated value as the number of data points in $\langle \Delta t \rangle$ from the same simulation increases. Corresponding measurement times in seconds for these data are given along the top axis. These data were generated from molecules at a concentration of $C = 34.60$ nM and diffusion rate of $D = 4.0$ $\mu m^2 s^{-1}$. (**C**) The MAL estimator for diffusion coefficient $\hat{D}_{\mathrm{MAL}}$ also accurately derives diffusion rates. (**D**) Estimated diffusion rates also converge to the true value as the number of data points (collected photons) increases. Error bars represent standard errors calculated via bootstrapping.

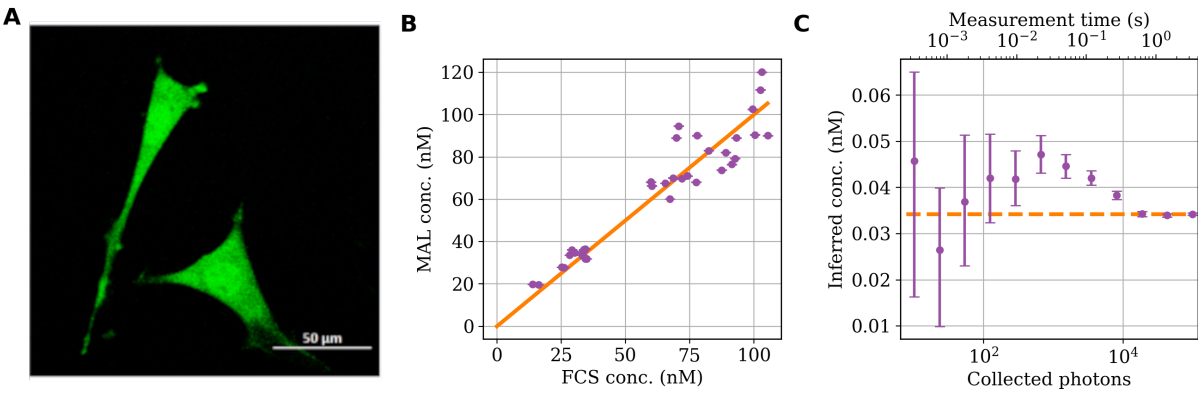

Figure 4: MAL accurately estimates concentrations in vivo (**A**) Mouse embryonic fibroblasts (MEF) expressing EGFP. Scale bar is 50 µ m (**B**) The MAL estimator for concentration $\hat{C}_{\mathrm{MAL}}$ is well correlated with estimates of concentration using standard FCS over 10 seconds of measurement (n=35). (**C**) MAL estimate of concentration over increasing number of data points demonstrates accurate inference can be made with only a few data points. Error bars represent standard errors calculated via bootstrapping.

which corresponds to hours of computational time. These long times limit the utility of this type of analysis.

In this study we introduced a new methodology for analysing FCS data, FCS-Maximum Approximate Likelihood (FCS-MAL), based upon an approximate likelihood model of an FCS experiment that draws inferences on the raw data. We applied carefully controlled approximations, namely Laplace's method, to solve the likelihood integral analytically. As such, FCS-MAL can both accurately quantify concentration and diffusion rates whilst maintaining the computational speed of traditional FCS analyses. Using FCS-MAL we can quantify molecular concentration and diffusion rates accurately with significantly less data. For a reasonably bright molecule fluorescing a few thousand times a seconds

MAL can successfully estimate these molecular parameters with as few as 10,000 collected photons, which corresponds to less than a 0.1 s of measurement time. This striking reduction in required measurement time for the same accuracy as standard FCS analyses requires no hardware modifications and is rapidly calculated on the fly from the raw photon arrivals data that is often deleted in most FCS experiments.

The methodology behind FCS-MAL is applicable to more complex models of molecular dynamics within an FCS experiment and presents a new framework to analysis processes such as anomalous and polydisperse diffusion (Höfling et al., 2011), photo-chemical processes like dark triplet states (Widengren et al., 1995), molecular interactions (Bacia et al., 2006), and DNA binding (Michelman-Ribeiro et al., 2009). Traditional FCS models for these processes already exist, yet remain difficult to implement as parameters such as triplet states and multiple diffusing species are hard to identify within the correlation curves is due to the low data density. Incorporating an anomalous parameter, α, would be possible by altering the variance in position to $\sigma_k = 2D(\Delta t_k)^\alpha$, followed by calculating the diffusion estimator alongside a new estimator for α. Therefore, our framework for approximating the likelihood opens up the possibility of robustly quantifying a greater range of molecular behaviours in vivo.

Acknowledgements

We thank the Bioimaging facility at the University of Manchester for microscopy support.

Data availability

Code for synthetic data generation and FCS analysis pipeline code has been made freely available via GitHub github.com/LoudonLab/FcsAnalysisPipeline.

References

Bacia K, Kim SA, Schwille P. Fluorescence cross-correlation spectroscopy in living cells. Nature Methods. 2006 Feb; 3(2):83–89. https://doi.org/10.1038/nmeth822, doi: 10.1038/nmeth822.

Bagnall J, Boddington C, Boyd J, Brignall R, Rowe W, Jones NA, Schmidt L, Spiller DG, White MR, Paszek P. Quantitative dynamic imaging of immune cell signalling using lentiviral gene transfer. Integr Biol (Camb). 2015; 7(6):713–25. https://www.ncbi.nlm.nih.gov/pubmed/25990200, doi: 10.1039/c5ib00067j, type: Journal Article.

Chen Y, Müller JD, So PTC, Gratton E. The Photon Counting Histogram in Fluorescence Fluctuation Spectroscopy. Biophysical Journal. 1999 Jul; 77(1):553–567. https://www.sciencedirect.com/science/article/pii/S0006349599769122, doi: 10.1016/S0006-3495(99)76912-2.

Efron B, Tibshirani R, Tibshirani RJ. An introduction to the bootstrap. Chapman & Hall/CRC Monographs on Statistics and Applied Probability, Philadelphia, PA: Chapman & Hall/CRC; 1994. http://dx.doi.org/10.1007/978-1-4899-4541-9, doi: 10.1007/978-1-4899-4541-9.

Elson EL, Magde D. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. Biopolymers. 1974; 13(1):1–27. https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.1974.360130102, doi: https://doi.org/10.1002/bip.1974.360130102, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bip.1974.360130102.

Haupts U, Maiti S, Schwille P, Webb WW. Dynamics of fluorescence fluctuations in green fluorescent protein observed by fluorescence correlation spectroscopy. Proceedings of the National Academy of Sciences. 1998; 95(23):13573–13578. https://www.pnas.org/doi/abs/10.1073/pnas.95.23.13573, doi: 10.1073/pnas.95.23.13573, _eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.95.23.13573.

Höfling F, Bamberg KU, Franosch T. Anomalous transport resolved in space and time by fluorescence correlation spectroscopy. Soft Matter. 2011; 7(4):1358–1363. http://dx.doi.org/10.1039/C0SM00718H, doi: 10.1039/C0SM00718H, publisher: The Royal Society of Chemistry.

Jazani S, Sgouralis I, Shafraz OM, Levitus M, Sivasankar S, Pressé S. An alternative framework for fluorescence correlation spectroscopy. Nature Communications. 2019 Aug; 10(1):3662. https://doi.org/10.1038/s41467-019-11574-2, doi: 10.1038/s41467-019-11574-2.

Kask P, Günther R, Axhausen P. Statistical accuracy in fluorescence fluctuation experiments. European Biophysics Journal. 1997 Jan; 25(3):163–169. https://doi.org/10.1007/s002490050028, doi: 10.1007/s002490050028.

Koch AA, Bagnall JS, Smyllie NJ, Begley N, Adamson AD, Fribourgh JL, Spiller DG, Meng QJ, Partch CL, Strimmer K, House TA, Hastings MH, Loudon AS. Quantification of protein abundance and interaction defines a mechanism for operation of the circadian clock. eLife. 2022 Mar; 11:e73976. https://doi.org/10.7554/eLife.73976, doi: 10.7554/eLife.73976, publisher: eLife Sciences Publications, Ltd.

Koppel DE. Statistical accuracy in fluorescence correlation spectroscopy. Phys Rev A. 1974 Dec; 10(6):1938–1945. https://link.aps.org/doi/10.1103/PhysRevA.10.1938, doi: 10.1103/PhysRevA.10.1938, publisher: American Physical Society.

Magde D, Elson E, Webb WW. Thermodynamic Fluctuations in a Reacting SystemMeasurement by Fluorescence Correlation Spectroscopy. Physical Review Letters. 1972 Sep; 29(11):705–708. https://link.aps.org/doi/10.1103/PhysRevLett.29.705, doi: 10.1103/PhysRevLett.29.705, number: 11.

Magidson V, Khodjakov A. Chapter 23 - Circumventing Photodamage in Live-Cell Microscopy. In: Sluder G, Wolf DE, editors. *Methods in Cell Biology*, vol. 114 Academic Press; 2013.p. 545–560. https://www.sciencedirect.com/science/article/pii/B9780124077614000233, doi: 10.1016/B978-0-12-407761-4.00023-3.

Michelman-Ribeiro A, Mazza D, Rosales T, Stasevich TJ, Boukari H, Rishi V, Vinson C, Knutson JR, McNally JG. Direct Measurement of Association and Dissociation Rates of DNA Binding in Live Cells by Fluorescence Correlation Spectroscopy. Biophysical Journal. 2009 Jul; 97(1):337–346. https://doi.org/10.1016/j.bpj.2009.04.027, doi: 10.1016/j.bpj.2009.04.027, publisher: Elsevier.

Murray JD. Laplace's method for integrals. In: *Asymptotic Analysis* Springer-verlag; 1984.p. 19–39.

Müller P, Python multiple-tau algorithm; 2012. https://pypi.python.org/pypi/multipletau/.

Politi AZ, Cai Y, Walther N, Hossain MJ, Koch B, Wachsmuth M, Ellenberg J. Quantitative mapping of fluorescently tagged cellular proteins using FCS-calibrated four-dimensional imaging. Nature Protocols. 2018 Jun; 13(6):1445–1464. https://doi.org/10.1038/nprot.2018.040, doi: 10.1038/nprot.2018.040.

Qian H. On the statistics of fluorescence correlation spectroscopy. Biophysical Chemistry. 1990 Oct; 38(1):49–57. https://www.sciencedirect.com/science/article/pii/030146229080039A, doi: 10.1016/0301-4622(90)80039-A.

Rigler R, Mets , Widengren J, Kask P. Fluorescence correlation spectroscopy with high count rate and low background: analysis of translational diffusion. European Biophysics Journal. 1993 Aug; 22(3):169–175. https://doi.org/10.1007/BF00185777, doi: 10.1007/BF00185777.

Sadaie W, Harada Y, Matsuda M, Aoki K. Quantitative In Vivo Fluorescence Cross-Correlation Analyses Highlight the Importance of Competitive Effects in the Regulation of Protein-Protein Interactions.

Molecular and Cellular Biology. 2014; 34(17):3272–3290. https://mcb.asm.org/content/34/17/3272, doi: 10.1128/MCB.00087-14, publisher: American Society for Microbiology Journals _eprint: https://mcb.asm.org/content/34/17/3272.full.pdf.

Saffarian S, Elson EL. Statistical Analysis of Fluorescence Correlation Spectroscopy: The Standard Deviation and Bias. Biophysical Journal. 2003; 84(3):2030–2042. https://www.sciencedirect.com/science/article/pii/S0006349503750115, doi: https://doi.org/10.1016/S0006-3495(03)75011-5.

Schwille P. Fluorescence correlation spectroscopy and its potential for intracellular applications. Cell Biochemistry and Biophysics. 2001 Jun; 34(3):383–408. https://doi.org/10.1385/CBB:34:3:383, doi: 10.1385/CBB:34:3:383.

Schwille P, Haupts U, Maiti S, Webb WW. Molecular Dynamics in Living Cells Observed by Fluorescence Correlation Spectroscopy with One- and Two-Photon Excitation. Biophysical Journal. 1999 Oct; 77(4):2251–2265. https://doi.org/10.1016/S0006-3495(99)77065-7, doi: 10.1016/S0006-3495(99)77065-7, publisher: Elsevier.

Schätzel K, Drewel M, Stimac S. Photon Correlation Measurements at Large Lag Times: Improving Statistical Accuracy. Journal of Modern Optics. 1988; 35(4):711–718. https://doi.org/10.1080/09500348814550731, doi: 10.1080/09500348814550731, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09500348814550731.

Stuart AM. Inverse problems: A Bayesian perspective. Acta Numerica. 2010; 19:451–559. https://www.cambridge.org/core/article/inverse-problems-a-bayesian-perspective/587A3A0D480A1A7C2B1B284BCEDF7E23, doi: 10.1017/S0962492910000061, edition: 2010/05/10 Publisher: Cambridge University Press.

Tavakoli M, Jazani S, Sgouralis I, Shafraz OM, Sivasankar S, Donaphon B, Levitus M, Pressé S. Pitching Single-Focus Confocal Data Analysis One Photon at a Time with Bayesian Nonparametrics. Physical Review X. 2020 Jan; 10(1):011021. https://link.aps.org/doi/10.1103/PhysRevX.10.011021, doi: 10.1103/PhysRevX.10.011021, publisher: American Physical Society.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020; 17:261–272. doi: 10.1038/s41592-019-0686-2.

Widengren J, Mets U, Rigler R. Fluorescence correlation spectroscopy of triplet states in solution: a theoretical and experimental study. The Journal of Physical Chemistry. 1995 Sep; 99(36):13368–13379. https://doi.org/10.1021/j100036a009, doi: 10.1021/j100036a009, publisher: American Chemical Society.

Yoo SH, Kojima S, Shimomura K, Koike N, Buhr ED, Furukawa T, Ko CH, Gloston G, Ayoub C, Nohara K, Reyes BA, Tsuchiya Y, Yoo OJ, Yagita K, Lee C, Chen Z, Yamazaki S, Green CB, Takahashi JS. Period2 3-UTR and microRNA-24 regulate circadian rhythms by repressing PERIOD2 protein accumulation. Proceedings of the National Academy of Sciences. 2017 Oct; 114(42):E8855–E8864. https://doi.org/10.1073/pnas.1706611114, doi: 10.1073/pnas.1706611114, publisher: Proceedings of the National Academy of Sciences.

Yoo SH, Shin Y, Lowrey PL, Kazuhiro S, H KC, Buhr Ethan D , Siepka Sandra M , Hong Hee-Kyung, Oh Won Jun, Yoo Ook Joon, Menaker Michael, Takahashi Joseph S . PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proceedings of the National Academy of Sciences. 2004 Apr; 101(15):5339–5346. https://doi.org/10.1073/pnas.0308709101, doi: 10.1073/pnas.0308709101, publisher: Proceedings of the National Academy of Sciences.

Yu L, Lei Y, Ma Y, Liu M, Zheng J, Dan D, Gao P. A Comprehensive Review of Fluorescence Correlation Spectroscopy. Frontiers in physics. 2021; 9. doi: 10.3389/fphy.2021.644450, type: Journal Article.

Zhang B, Zerubia J, Olivo-Marin JC. Gaussian approximations of fluorescence microscope point-spread function models. Applied Optics. 2007 Apr; 46(10):1819–1829. https://opg.optica.org/ao/abstract.cfm?URI=ao-46-10-1819, doi: 10.1364/AO.46.001819, publisher: Optica Publishing Group.

Appendix

Synthetic data generation

Monte Carlo simulations of FCS experiments were performed in the Python 3 programming language by modelling molecules moving within a defined volume $V_{\text{box}} = L_{xy}^2 L_z$ as a random walk according to the same scheme as Jazani et al. (2019). The code of the main loop of the random walk algorithm code was compiled via Cython to improve computational performance. Initial positions within the volume are drawn from a uniform distribution as

$$x_{n,0}, y_{n,0} \sim \textbf{Uniform}(-\frac{L_{xy}}{2}, \frac{L_{xy}}{2}), \quad z_{n,0} \sim \textbf{Uniform}(-\frac{L_z}{2}, \frac{L_z}{2}). \tag{12}$$

Molecular positions in each direction, x, y, and z are updated by adding random variables drawn from a normal distribution of mean zero and variance $2D\delta t$ according to

$$\begin{aligned} x_{n,k+1} &\sim x_{n,k} + \textbf{Normal}(0, 2D\delta t) \\ y_{n,k+1} &\sim y_{n,k} + \textbf{Normal}(0, 2D\delta t) \\ z_{n,k+1} &\sim z_{n,k} + \textbf{Normal}(0, 2D\delta t). \end{aligned} \tag{13}$$

If the proposed new position takes is beyond the bounding box, V_{box}, then the position of the particle is reassigned to the other side, i.e. if the particle is now to the right of the boundary $\frac{L_{xy}}{2}$ in the positive x direction then $x_{n,k+1} \rightarrow x_{n,k+1} - L_{xy}$. Finally, the number of photons emitted in the interval δt is determined from the distribution

$$\gamma_k \sim \textbf{Poisson}(\lambda_k), \tag{14}$$

where the average number of photons emitted from N molecules at their current positions is the sum

$$\lambda_k = \mu_m \delta t \sum_{n=1}^{N} \text{PSF}(x_{n,k}, y_{n,k}, z_{n,k}). \tag{15}$$

Here, the point spread function (PSF) models the optics of the system and computes the normalised probability of a photon being emitted at the current position of the molecule (Chen et al., 1999). The molecular brightness per unit time μ_m encodes the probability of excitation by the laser, quantum yield of the fluorophore, and likelihood of detection, hence when multiplying the sum of PSFs produces the average rate of detected photons. In this study we exclusively use a 3D Gaussian PSF defined as

$$\text{PSF}_{3\text{G}}(x_{n,k}, y_{n,k}, z_{n,k}) = \exp\left(-2\left(\frac{x_{n,k}^2 + y_{n,k}^2}{w_{xy}^2} + \frac{z_{n,k}^2}{w_z^2}\right)\right) \tag{16}$$

where w_{xy} and w_z are the radial and axial radii of the confocal volume, and $w_z > w_{xy}$. Finally, the number of photon emissions is a random variable within each time interval δt drawn from the Poisson distribution defined in Equation 14. In practice the sampling time step δt is chosen to be small enough such that $\gamma_k \leq 1$ to avoid over sampling. We subsequently calculate the time intervals between photons, $\Delta t_k = t_{k+1} - t_k$, from these data.

Detailed calculation of approximate likelihood distribution

Here we infer the quantity and diffusion of fluorescent molecules within a FCS experiment. We begin with a similar approach as Tavakoli et al. (2020) by calculating the likelihood of observing the time intervals between detected photons $\boldsymbol{\Delta t}_k$ for the set of parameters $\boldsymbol{\theta}$; $P(\boldsymbol{\Delta t}|\boldsymbol{\theta})$. In general $\boldsymbol{\theta}$ is a vector of all time steps k

$$\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_K), \tag{17}$$

$$\theta_k = \{\mu_m^{(k)}, D_k, N_k, w_{xy}^{(k)}, w_z^{(k)}\}; \quad k = 1, 2, \ldots, K \tag{18}$$

which we will simplify by assuming the parameters are constant over the experiment such that

$$\theta_k = \theta = \{\mu_m, D, N, w_{xy}, w_z\} \tag{19}$$

for all k and

$$P(\boldsymbol{\Delta t}|\underline{\theta}) = P(\boldsymbol{\Delta t}|\theta). \tag{20}$$

The difficulty in moving forward is that the unobserved intermediate positions of each of the N particles, $\underline{\underline{\boldsymbol{r}}}$, have been marginalised out. Therefore, let's expand the likelihood as a marginal integration over the possible values of $\underline{\underline{\boldsymbol{r}}}$. To be clear let's define the possible positions over 3 dimensions, K time steps and N particles as

$$\begin{aligned}
&\text{Over dimensions } x, y, z: \ \boldsymbol{r}_{n,k} = (x_{n,k}, y_{n,k}, z_{n,k}) \\
&\text{Over particles } N: \ \underline{\boldsymbol{r}}_k = (\underline{\boldsymbol{r}}_1, \underline{\boldsymbol{r}}_2, \ldots, \underline{\boldsymbol{r}}_N); \quad n = 1, 2, \ldots, N \\
&\text{Over time steps } K: \ \underline{\underline{\boldsymbol{r}}} = (\boldsymbol{r}_{n,1}, \boldsymbol{r}_{n,2}, \ldots, \boldsymbol{r}_{n,K}); \quad k = 1, 2, \ldots, K.
\end{aligned} \tag{21}$$

Using Total Probability $P(A) = \int_B P(A|B)P(B)dB$ and Conditional Probability $P(A, B) = P(A|B)P(B)$, the likelihood is the marginal integral

$$\begin{aligned}
P(\boldsymbol{\Delta t}|\theta) &= \frac{P(\boldsymbol{\Delta t}, \theta)}{P(\theta)} \\
&= \int_{\underline{\underline{\boldsymbol{r}}}} \frac{P(\boldsymbol{\Delta t}, \theta|\underline{\underline{\boldsymbol{r}}})P(\underline{\underline{\boldsymbol{r}}})}{P(\theta)} d\underline{\underline{\boldsymbol{r}}} \\
&= \int_{\underline{\underline{\boldsymbol{r}}}} \frac{P(\boldsymbol{\Delta t}, \underline{\underline{\boldsymbol{r}}}, \theta)}{P(\theta)} d\underline{\underline{\boldsymbol{r}}} \\
&= \int_{\underline{\underline{\boldsymbol{r}}}} \frac{P(\boldsymbol{\Delta t}|\underline{\underline{\boldsymbol{r}}}, \theta)P(\underline{\underline{\boldsymbol{r}}}, \theta)}{P(\theta)} d\underline{\underline{\boldsymbol{r}}} \\
&= \int_{\underline{\underline{\boldsymbol{r}}}} \underbrace{P(\boldsymbol{\Delta t}|\underline{\underline{\boldsymbol{r}}}, \theta)}_{\substack{\text{Photon emission} \\ \text{and detection}}} \underbrace{P(\underline{\underline{\boldsymbol{r}}}|\theta)}_{\text{Diffusion}} d\underline{\underline{\boldsymbol{r}}}
\end{aligned} \tag{22}$$

Inside the integrand, the first term describes how times between arrivals depends on the positions of the molecules and parameters θ, whereas the second terms defines how molecular positions are distributed over time and space. We may expand $P(\boldsymbol{\Delta t}|\underline{\underline{\boldsymbol{r}}})$ over each time step using conditional probability

$$\begin{aligned}
P(\boldsymbol{\Delta t}|\underline{\underline{\boldsymbol{r}}}, \theta) =& \frac{P(\Delta t_{K-1}, \Delta t_{K-2}, \ldots, \Delta t_1, \underline{\underline{\boldsymbol{r}}}, \theta)}{P(\underline{\underline{\boldsymbol{r}}}, \theta)} \\
=& \frac{P(\Delta t_{K-1}|\Delta t_{K-2}, \ldots, \Delta t_1, \underline{\underline{\boldsymbol{r}}}, \theta)P(\Delta t_{K-2}, \ldots, \Delta t_1, \underline{\underline{\boldsymbol{r}}}, \theta)}{P(\underline{\underline{\boldsymbol{r}}}, \theta)} \\
=& \frac{P(\Delta t_{K-1}|\Delta t_{K-2}, \ldots, \Delta t_1, \underline{\underline{\boldsymbol{r}}}, \theta)P(\Delta t_{K-2}|\Delta t_{K-3}, \ldots, \Delta t_1, \underline{\underline{\boldsymbol{r}}}, \theta)}{P(\underline{\underline{\boldsymbol{r}}}, \theta)} \ldots \\
&\times \frac{P(\Delta t_2|\Delta t_1, \underline{\underline{\boldsymbol{r}}}, \theta)P(\Delta t_1|\underline{\underline{\boldsymbol{r}}}, \theta)P(\underline{\underline{\boldsymbol{r}}}, \theta)}{P(\underline{\underline{\boldsymbol{r}}}, \theta)} \\
=& P(\Delta t_{K-1}|\Delta t_{K-2}, \ldots, \Delta t_1, \underline{\underline{\boldsymbol{r}}}, \theta)P(\Delta t_{K-2}|\Delta t_{K-3}, \ldots, \Delta t_1, \underline{\underline{\boldsymbol{r}}}, \theta) \ldots \\
&\times P(\Delta t_2|\Delta t_1, \underline{\underline{\boldsymbol{r}}}, \theta)P(\Delta t_1|\underline{\underline{\boldsymbol{r}}}, \theta).
\end{aligned} \tag{23}$$

This can then be simplified by assuming that photon emissions are memoryless (Markovian), such that the current probability of a photon being emitted is only dependent on the current molecular positions and emission parameters. Time between emissions in the interval $k \to k + 1$ (Δt_k) only depend on the current positions of the particles $\underline{\boldsymbol{r}}_k$ and are non-interacting. i.e. time between photon emissions is a Markovian process

$$P(\Delta t_k|\Delta t_{k-1}, \ldots, \Delta t_1, \underline{\underline{\boldsymbol{r}}}, \theta) = P(\Delta t_k|\underline{\boldsymbol{r}}_k, \theta). \tag{24}$$

Thus, Equation 23 becomes

$$P(\boldsymbol{\Delta t}|\underline{\underline{\boldsymbol{r}}},\theta) = \prod_{k=1}^{K-1} P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta). \tag{25}$$

Again using conditional probability, the diffusive term of Equation 22, $P(\underline{\underline{\boldsymbol{r}}}|\theta)$, may be expanded as

$$\begin{aligned} P(\underline{\underline{\boldsymbol{r}}}|\theta) &= \frac{P(\underline{\boldsymbol{r}}_K,\underline{\boldsymbol{r}}_{K-1},\ldots,\underline{\boldsymbol{r}}_1,\theta)}{P(\theta)} \\ &= \frac{P(\underline{\boldsymbol{r}}_K|\underline{\boldsymbol{r}}_{K-1},\ldots,\underline{\boldsymbol{r}}_1,\theta)P(\underline{\boldsymbol{r}}_{K-1}|\underline{\boldsymbol{r}}_{K-2},\ldots,\underline{\boldsymbol{r}}_1,\theta)\ldots P(\underline{\boldsymbol{r}}_2|\underline{\boldsymbol{r}}_1,\theta)P(\underline{\boldsymbol{r}}_1|\theta)P(\theta)}{P(\theta)} \\ &= P(\underline{\boldsymbol{r}}_K|\underline{\boldsymbol{r}}_{K-1},\ldots,\underline{\boldsymbol{r}}_1,\theta)P(\underline{\boldsymbol{r}}_{K-1}|\underline{\boldsymbol{r}}_{K-2},\ldots,\underline{\boldsymbol{r}}_1,\theta)\ldots P(\underline{\boldsymbol{r}}_2|\underline{\boldsymbol{r}}_1,\theta)P(\underline{\boldsymbol{r}}_1|\theta). \end{aligned} \tag{26}$$

As before, we now assume that the diffusive process is Markovian, only dependent on the previous molecular positions and not the full positional history of the molecules

$$P(\underline{\boldsymbol{r}}_{k+1}|\underline{\boldsymbol{r}}_k,\ldots,\underline{\boldsymbol{r}}_1,\theta) = P(\underline{\boldsymbol{r}}_{k+1}|\underline{\boldsymbol{r}}_k,\theta). \tag{27}$$

Hence, $P(\underline{\underline{\boldsymbol{r}}}|\theta)$ is the product

$$P(\underline{\underline{\boldsymbol{r}}}|\theta) = \left(\prod_{k=1}^{K-1} P(\underline{\boldsymbol{r}}_{k+1}|\underline{\boldsymbol{r}}_k,\theta)\right) P(\underline{\boldsymbol{r}}_1|\theta). \tag{28}$$

In summary, we have now defined the likelihood in Equation 22 as

$$P(\boldsymbol{\Delta t}|\theta) = \int_{\underline{\underline{\boldsymbol{r}}}} \left(\prod_{k=1}^{K-1} P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta)P(\underline{\boldsymbol{r}}_{k+1}|\underline{\boldsymbol{r}}_k,\theta)\right) P(\boldsymbol{r}_1|\theta)d\underline{\underline{\boldsymbol{r}}}. \tag{29}$$

We define the integration measure $d\underline{\underline{\boldsymbol{r}}}$ as simply a product over times, K, particles, N, and dimensions, x, y, z

$$d\underline{\underline{\boldsymbol{r}}} = \prod_{n=1}^{N}\prod_{k=1}^{K} dx_{n,k}dy_{n,k}dz_{n,k}. \tag{30}$$

Therefore, the likelihood from Equation 29 becomes

$$P(\boldsymbol{\Delta t}|\theta) = \int_{\underline{\underline{\boldsymbol{r}}}} \left(\prod_{k=1}^{K-1} P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta)P(\underline{\boldsymbol{r}}_{k+1}|\underline{\boldsymbol{r}}_k,\theta)\right) P(\underline{\boldsymbol{r}}_1|\theta) \left(\prod_{n=1}^{N}\prod_{k=1}^{K} dx_{n,k}dy_{n,k}dz_{n,k}\right), \tag{31}$$

which represents the furthest we can go without specifying the optics of the system or the model of diffusion. Specifying these requires us to define the distributions for $P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta)$ and $P(\boldsymbol{r}_{k+1}|\underline{\boldsymbol{r}}_k,\theta)$.

3D normal diffusion in a 3D Gaussian Point Spread Function with multiple particles

A straightforward model for 3D molecular movement within a solution is that of molecules diffusing according to normal diffusion. Additionally, the optics of many confocal microscopes are well described by a 3D Gaussian point spread function. Here, photon emissions from the fluorescent molecules are distributed by an exponential distribution $P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta)$ conditioned on molecular positions and emission parameters

$$P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta) = \textbf{Exponential}(\mu_k) \tag{32}$$

such that with the same 3D Gaussian PSF (3G) defined in Equation 16 we get

$$\mu_k = \mu_m \sum_{n=1}^{N} \mathrm{PSF}_{3\mathrm{G}}(\underline{\boldsymbol{r}}_k). \tag{33}$$

Therefore,

$$\begin{aligned} P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta) &= \mu_k \exp\left(-\mu_k \Delta t_k\right) \\ &= \mu_m \sum_{n=1}^{N} \exp\left[-2\left(\frac{x_{n,k}^2+y_{n,k}^2}{w_{xy}^2}+\frac{z_{n,k}^2}{w_z^2}\right)\right) - \Delta t_k \mu_m \sum_{n'=1}^{N} \exp\left(-2\left(\frac{x_{n',k}^2+y_{n',k}^2}{w_{xy}^2}+\frac{z_{n',k}^2}{w_z^2}\right)\right] \end{aligned} \tag{34}$$

Laplace approximation for $P(\Delta t|\theta)$

Given the double exponential terms it is clear we cannot calculate $P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta)$ in closed form as the integral over $\underline{\underline{\boldsymbol{r}}}$ is non-trivial. Fortunately, as the form of Equation 34 is highly peaked and symmetrical we may approximate it using Laplace's method as a Gaussian function centered on the maxima at $\underline{\underline{\boldsymbol{r}}} = \underline{0}$. This approximation is valid when $\mu_m \Delta t_k < 1$ over all k, and asymptotically approaches the true value of $P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta)$ as $\mu_m \Delta t_k \to 0$. We apply this method to each term in the sum

$$P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta) = \mu_m \sum_{n=1}^{N} e^{h_n(\underline{\boldsymbol{r}}_k)}. \tag{35}$$

Key to Laplace's method is the construction of a Gaussian function via a Taylor expansion to second order of the exponent $h_n(\boldsymbol{r})$ around the maxima. In the case of $\mu_m \Delta t_k < 1$ the maxima occurs at the centre of the confocal volume $\boldsymbol{r}_0 = 0$. To simplify notation in further calculations let's define $M_k = \mu_m \Delta t_k$, $a_d = \frac{2}{w_d^2}$ for dimension $d = xy$ or $d = z$, and $C_{d,k} = a_d(1 - M_k)$. Omitting the time interval k subscript for now gives

$$h_n(\underline{\boldsymbol{r}}) = -a_{xy}x_n^2 - a_{xy}y_n^2 - a_z z_n^2 - M \sum_{n'=1}^{N} \exp\left(-a_{xy}x_{n'}^2 - a_{xy}y_{n'}^2 - a_z z_{n'}^2\right) \tag{36}$$

The Taylor expansion for h_n is

$$h_n(\boldsymbol{p}) \approx h_n(\boldsymbol{p}_0) + \sum_{j=1}^{3N} \frac{\partial h_n(\boldsymbol{p}_0)}{\partial p_j}(p_j - p_{j,0}) + \frac{1}{2}\sum_{i=1}^{3N}\sum_{j=1}^{3N} \frac{\partial^2 h_n(\boldsymbol{p}_0)}{\partial p_i \partial p_j}(p_i - p_{i,0})(p_j - p_{j,0}) + O((\boldsymbol{p} - \boldsymbol{p}_0)^3), \tag{37}$$

where for ease we have redefined the vector $\boldsymbol{r}$ with N components into a flat vector $\boldsymbol{p}$ with $3N$ element. The first N indices are all the x's, $N+1$ to $2N$ are the y's, and finally $2N+1$ to $3N$ are the z's such that $\boldsymbol{p} = (x_1, x_2, \ldots, x_N, y_1, y_2, \ldots, y_N, z_1, z_2, \ldots, z_N)$. As the function we are approximating is symmetrical and is stationary at the maxima $\boldsymbol{p}_0$ we get $\frac{h_n(\boldsymbol{p}_0)}{\partial p_j} = 0$. The non-zero terms are also readily computed, with $h_n(\boldsymbol{p}_0) = -MN$ and

$$\frac{\partial^2 h_n^2(\boldsymbol{p}_0)}{\partial p_i \partial p_j} = \begin{cases} -2a_d(1-M) & i = j = n \\ 2a_d M & (i = j) \neq n \\ 0 & \text{otherwise} \end{cases} \tag{38}$$

where $d = xy$ if $n \leq 2N$ otherwise $d = z$. As $M < 1$ this ensures that the curvature of the function is negative at the maxima. Putting all these terms together gives

$$h_n(\boldsymbol{p}) \approx -MN - a_{xy}x_n^2 - a_{xy}y_n^2 - a_z z_n^2 + M\left(\sum_{n'=1}^{N} a_{xy}x_{n'}^2 + a_{xy}y_{n'}^2 + a_z z_{n'}^2\right), \tag{39}$$

All together this gives

$$P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta) \approx \mu_m \sum_{n=1}^{N} \exp\left(-MN - a_{xy}x_n^2 - a_{xy}y_n^2 - a_z z_n^2 + M\left(\sum_{n'=1}^{N} a_{xy}x_{n'}^2 + a_{xy}y_{n'}^2 + a_z z_{n'}^2\right)\right). \tag{40}$$

This can be factored using the original definition of $\mathrm{PSF}_{3\mathrm{G}}(\boldsymbol{r}_{n,k})$ as

$$P(\Delta t_k|\underline{\boldsymbol{r}}_k,\theta) \approx \mu_m \left(\sum_{n=1}^{N} \mathrm{PSF}_{3\mathrm{G}}(\boldsymbol{r}_{n,k})\right)\left(e^N \prod_{n=1}^{N} \mathrm{PSF}_{3\mathrm{G}}(\boldsymbol{r}_{n,k})\right)^{-\mu_m \Delta t_k}. \tag{41}$$

Diffusion and molecular positions distributions

The term $P(\boldsymbol{r}_k|\underline{\boldsymbol{r}}_{k-1},\theta)$ describes how positions update in 3 dimensions over each time interval. In this study we focus on normal 3D diffusion with diffusion coefficient D. Particle positions at a later time $k+1$ are a multivariate normal distribution with means conditioned on positions at the previous time k for $K > 1$

$$P(\underline{\boldsymbol{r}}_{k+1}|\underline{\boldsymbol{r}}_k,\theta) = \frac{1}{\sqrt{8\pi\sigma_k^3}} \exp\left(-\frac{1}{\sigma_k}\left[(x_{n,k+1}-x_{n,k})^2 + (y_{n,k+1}-y_{n,k})^2 + (z_{n,k+1}-z_{n,k})^2\right]\right) \tag{42}$$

for $\sigma_k = 2D\Delta t_k$. However, this only describes how the particle moves and not its initial spatial distribution. Therefore, we will assume that particles are uniformly distributed on a line of length L. Particle positions are initially Uniformly distributed in each orthogonal dimension at time $k = 1$

$$P(\underline{\boldsymbol{r}}_1|\theta) = \frac{1}{L_{xy}^2 L_z}. \tag{43}$$

Thus, the full set of positions over time is

$$\begin{aligned} P(\underline{\underline{\boldsymbol{r}}}|\theta) &= \left(\prod_{k=1}^{K-1} P(\underline{\boldsymbol{r}}_{k+1}|\underline{\boldsymbol{r}}_k,\theta)\right) P(\underline{\boldsymbol{r}}_1|\theta) \\ &= \frac{1}{L_{xy}^2 L_z} \prod_{k=1}^{K-1}\prod_{n=1}^{N} \frac{1}{(2\pi\sigma_k)^{\frac{3}{2}}} \exp\left(-\frac{1}{\sigma_k}\left[(x_{n,k+1}-x_{n,k})^2 + (y_{n,k+1}-y_{n,k})^2 + (z_{n,k+1}-z_{n,k})^2\right]\right) \\ &= \frac{1}{V_{\text{box}}} \prod_{k=1}^{K-1} \frac{1}{(2\pi\sigma_k)^{\frac{3N}{2}}} \prod_{n=1}^{N} \exp\left(-\frac{1}{\sigma_k}\left[(x_{n,k+1}-x_{n,k})^2 + (y_{n,k+1}-y_{n,k})^2 + (z_{n,k+1}-z_{n,k})^2\right]\right) \end{aligned} \tag{44}$$

Calculating the integral

We now have all the necessary components to complete the computation of the likelihood. Explicitly the likelihood integral is

$$\begin{aligned} P(\boldsymbol{\Delta t}|\theta) = \frac{\mu_m^{K-1}}{V_{\text{box}}} \int_{\underline{\underline{\boldsymbol{r}}}} \prod_{k=1}^{K-1} \frac{1}{(2\pi\sigma_k)^{\frac{3N}{2}}} \Bigg(\sum_{n=1}^{N} \exp\Big[&-\mu_m N\Delta t_k - a_{xy}x_{n,k+1}^2 - a_{xy}y_{n,k+1}^2 - a_z z_{n,k+1}^2 \\ &+ M\Big(\sum_{n'=1}^{N} a_{xy}x_{n',k+1}^2 + a_{xy}y_{n',k+1}^2 + a_z z_{n',k+1}^2\Big)\Big] \\ &\times \exp\Big[-\frac{1}{\sigma_k}\left[(x_{n,k+1}-x_{n,k})^2 + (y_{n,k+1}-y_{n,k})^2 + (z_{n,k+1}-z_{n,k})^2\right]\Big] d\underline{\underline{\boldsymbol{r}}}. \end{aligned} \tag{45}$$

Some terms can be taken outside the integral and the directions, x, y, and z can be separated out as

$$\begin{aligned} P(\boldsymbol{\Delta t}|\theta) = \frac{\mu_m^{K-1}}{V_{\text{box}}} \left(\prod_{k=1}^{K-1} \frac{e^{-\mu_m N\Delta t_k}}{(2\pi\sigma_k)^{\frac{3N}{2}}}\right) &\underbrace{\int_{\underline{\underline{x}}} \prod_{k=1}^{K-1} \left(\sum_{n=1}^{N} \exp\left(-a_{xy}x_{n,k}^2 + \mu_m\Delta t_k \sum_{n'=1}^{N} a_{xy}x_{n',k}^2\right) \exp\left(-\frac{(x_{n,k+1}-x_{n,k})^2}{\sigma_k}\right)\right) d\underline{\underline{x}}}_{P_{N,K}^{(x)}} \\ &\times \underbrace{\int_{\underline{\underline{y}}} \prod_{k=1}^{K-1} \left(\sum_{n=1}^{N} \exp\left(-a_{xy}y_{n,k}^2 + \mu_m\Delta t_k \sum_{n'=1}^{N} a_{xy}y_{n',k}^2\right) \exp\left(-\frac{(y_{n,k+1}-y_{n,k})^2}{\sigma_k}\right)\right) d\underline{\underline{y}}}_{P_{N,K}^{(y)}} \\ &\times \underbrace{\int_{\underline{\underline{z}}} \prod_{k=1}^{K-1} \left(\sum_{n=1}^{N} \exp\left(-a_z z_{n,k}^2 + \mu_m\Delta t_k \sum_{n'=1}^{N} a_z z_{n',k}^2\right) \exp\left(-\frac{(z_{n,k+1}-z_{n,k})^2}{\sigma_k}\right)\right) d\underline{\underline{z}}}_{P_{N,K}^{(z)}}. \end{aligned} \tag{46}$$

Here each of the integrals, $P_{N,K}^{(x)}$, $P_{N,K}^{(y)}$, and, $P_{N,K}^{(z)}$ over x, y, and z respectively are independent of one another. Thus, we can factor the integrals

$$P(\boldsymbol{\Delta t}|\theta) = \frac{\mu_m^{K-1}}{V_{\text{box}}} \left(\prod_{k=1}^{K-1} \frac{e^{-\mu_m N \Delta t_k}}{(2\pi\sigma_k)^{\frac{3N}{2}}} \right) P_{N,K}^{(x)} P_{N,K}^{(y)} P_{N,K}^{(z)}. \tag{47}$$

To calculate the integrals we initially focus on $P_{N,K}^{(x)}$. Using the definitions $C_{d,k} = a_d(1 - M_k)$ where d is either xy or z, and $M_k = \mu_m \Delta t_k$ we can rearrange $P_{N,K}^{(x)}$ as

$$\begin{aligned} P_{N,K}^{(x)} &= \int_{\underline{\underline{x}}} \prod_{k=1}^{K-1} \left(\sum_{n=1}^{N} \exp\left(-C_{xy,k} x_{n,k}^2 + a_{xy} M_k \sum_{n' \neq n=1}^{N} x_{n',k}^2 \right) \left(\prod_{n'=1}^{N} \exp\left(-\frac{\Delta x_{n',k}^2}{\sigma_k} \right) \right) \right) d\underline{\underline{x}} \\ &= \int_{\underline{\underline{x}}_{-1}} \prod_{k=2}^{K-1} \ldots d\underline{\underline{x}}_{-1} \int_{\underline{x}_1} \sum_{n=1}^{N} \exp\left(-C_{xy,1} x_{n,1}^2 + a_{xy} M_1 \sum_{n' \neq n=1}^{N} x_{n',1}^2 \right) \left(\prod_{n'=1}^{N} \exp\left(-\frac{\Delta x_{n',1}^2}{\sigma_1} \right) \right) d\underline{x}_1 \\ &= \int_{\underline{\underline{x}}_{-1}} \prod_{k=2}^{K-1} \ldots d\underline{\underline{x}}_{-1} \sum_{n=1}^{N} \underbrace{\int_{\underline{x}_1} \exp\left(-C_{xy,1} x_{n,1}^2 + a_{xy} M_1 \sum_{n' \neq n=1}^{N} x_{n',1}^2 \right) \left(\prod_{n'=1}^{N} \exp\left(-\frac{\Delta x_{n',1}^2}{\sigma_1} \right) \right) d\underline{x}_1}_{I_{n,1}} . \end{aligned} \tag{48}$$

Succinctly, this may be written as

$$P_{N,K}^{(x)} = \prod_{k=1}^{K-1} \left(\sum_{n=1}^{N} I_{n,k} \right). \tag{49}$$

Practically this means that each $P_{N,K}^{(x)}$ is a sum over all possible combinations of integrals over molecules n at time steps k. This is such that N chains of integrals follow the same particle over all time steps, whereas others swap over from as little as one time to N times. The full set of N^{K-1} sets of integrals for $P_{N,K}^{(x)}$ can be written as

$$\begin{aligned} P_{N,K}^{(x)} = {} & I_{1,K} I_{1,K-1} \ldots I_{1,1} + I_{2,K} I_{2,K-1} \ldots I_{2,1} + \cdots + I_{N,K} I_{N,K-1} \ldots I_{N,1} \\ & + I_{1,K} I_{2,K-1} \ldots I_{1,1} + I_{2,K} I_{1,K-1} \ldots I_{2,1} + \cdots + I_{N,K} I_{1,K-1} \ldots I_{N,1} \\ & \vdots \\ & + I_{1,K} I_{2,K-1} \ldots I_{N,1} + I_{2,K} I_{1,K-1} \ldots I_{N,1} + \cdots + I_{N,K} I_{N-1,K-1} \ldots I_{1,1} \\ & \vdots \\ & + I_{1,K} I_{1,K-1} \ldots I_{N,1} + I_{2,K} I_{2,K-1} \ldots I_{N,1} + \cdots + I_{N-1,K} I_{N-1,K-1} \ldots I_{N,1} \end{aligned} \tag{50}$$

For now, let's consider the integral for the particle n at the first time step $k = 1$

$$\begin{aligned} I_{n,1} &= \int_{\underline{x}_1} \exp\left(-C_{xy,1} x_{n,1}^2 + a_{xy} M_1 \sum_{n' \neq n=1}^{N} x_{n',1}^2 \right) \left(\prod_{n'=1}^{N} \exp\left(-\frac{\Delta x_{n',1}^2}{\sigma_1} \right) \right) d\underline{x}_1 \\ &= \int_{\underline{x}_1} \exp\left(-C_{xy,1} x_{n,1}^2 \right) \left(\prod_{n' \neq n=1}^{N} \exp\left(a_{xy} M_1 x_{n',1}^2 \right) \right) \left(\prod_{n'=1}^{N} \exp\left(-\frac{\Delta x_{n',1}^2}{\sigma_1} \right) \right) d\underline{x}_1 \\ &= \left(\prod_{n' \neq n=1}^{N} \int_{-\infty}^{\infty} \exp\left(-\frac{\Delta x_{n',1}^2}{\sigma_k} + a_{xy} M_1 x_{n',1}^2 \right) dx_{n',1} \right) \int_{-\infty}^{\infty} \exp\left(-C_{xy,1} x_{n,1}^2 - \frac{\Delta x_{n,1}^2}{\sigma_1} \right) dx_{n,1} \end{aligned} \tag{51}$$

Where we have taken the domain of $\underline{x}_1$ for all particles to be over all space from $-\infty$ to ∞. This approximation holds well in our model as the volume of the container, i.e. a cell, V_{box} is much larger than the confocal volume, V_{eff}, within which we are measuring. Let's expand $\Delta x_{n,1}^2 = x_{n,2}^2 - 2x_{n,2}x_{n,1} + x_{n,1}^2$ as well as define $\beta_{xy,1} = \frac{1}{\sigma_1} + C_{xy,1}$ and $\beta'_{xy,1} = \frac{1}{\sigma_1} - a_{xy} M_1$ to get

$$I_{n,1} = \left(\prod_{n' \neq n=1}^{N} \int_{-\infty}^{\infty} \exp\left(-\beta'_{xy,1} x_{n',1}^2 + \frac{2x_{n',2}}{\sigma_1} x_{n',1} - \frac{x_{n',2}^2}{\sigma_1} \right) dx_{n',1} \right) \int_{-\infty}^{\infty} \exp\left(-\beta_{xy,1} x_{n,1}^2 - \frac{2x_{n,2}}{\sigma_1} x_{n,1} - \frac{x_{n,2}^2}{\sigma_1} \right) dx_{n,1}. \tag{52}$$

This can then finally be integrated using the standard Gaussian integral

$$\int_{-\infty}^{\infty} e^{-ax^2+bx+c} dx = \sqrt{\frac{\pi}{a}} e^{\frac{b^2}{4a}} e^c, \tag{53}$$

to integrate $I_{n,1}$ as

$$\begin{aligned} I_{n,1} &= \left(\prod_{n' \neq n=1}^{N} \sqrt{\frac{\pi}{\beta'_{xy,1}}} \exp\left(-\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_1^2 \beta'_{xy,1}} \right) x_{n',2}^2 \right) \right) \sqrt{\frac{\pi}{\beta_{xy,1}}} \exp\left(-\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_1^2 \beta_{xy,1}} \right) x_{n,2}^2 \right) \\ &= \underbrace{\sqrt{\frac{\pi}{\beta_{xy,1}}} \left(\frac{\pi}{\beta'_{xy,1}} \right)^{\frac{N-1}{2}}}_{\text{Constants}} \underbrace{\left(\prod_{n' \neq n=1}^{N} \exp\left(-\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_1^2 \beta'_{xy,1}} \right) x_{n',2}^2 \right) \right) \exp\left(-\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_1^2 \beta_{xy,1}} \right) x_{n,2}^2 \right)}_{\text{Carry forward terms}} \end{aligned} \tag{54}$$

This demonstrates that we can integrate the first part of the full likelihood integral. Despite the complicated nature of this intermediate result it has a relatively simple interpretation that for every time step, k, we pick up constants and carry forward exponents into the next set of $k+1$ integrals. The $\beta_{xy,1}$ and $\beta'_{xy,1}$ terms are more complex as they encompass the constants μ_m, a_{xy}, and a_z as well as the data point Δt_1. To complete the full likelihood integral means integrating over all of the positions for all time steps. Further complication arises in the full set of integrals as mixing of particle identities between time steps gives different results. When the particle identity is the same we get

$$\begin{aligned} I_{n,2} I_{n,1} &= \left(\prod_{n' \neq n=1}^{N} \int_{-\infty}^{\infty} \exp\left(-\beta'_{xy,2} x_{n',2}^2 + \frac{2x_{n',3}}{\sigma_2} x_{n',2} - \frac{x_{n',3}^2}{\sigma_2} \right) dx_{n',2} \right) \int_{-\infty}^{\infty} \exp\left(-\beta_{xy,2} x_{n,2}^2 - \frac{2x_{n,3}}{\sigma_2} x_{n,2} - \frac{x_{n,3}^2}{\sigma_2} \right) dx_{n,2} \\ &= \left(\prod_{n' \neq n=1}^{N} \sqrt{\frac{\pi}{\beta'_{xy,2}}} \exp\left(-\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_2^2 \beta'_{xy,2}} \right) x_{n',3}^2 \right) \right) \sqrt{\frac{\pi}{\beta_{xy,1}}} \exp\left(-\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_2^2 \beta_{xy,2}} \right) x_{n,3}^2 \right) \\ &= \sqrt{\frac{\pi}{\beta_{xy,2}}} \left(\frac{\pi}{\beta'_{xy,2}} \right)^{\frac{N-1}{2}} \left(\prod_{n' \neq n=1}^{N} \exp\left(-\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_2^2 \beta'_{xy,2}} \right) x_{n',3}^2 \right) \right) \exp\left(-\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_2^2 \beta_{xy,2}} \right) x_{n,3}^2 \right), \end{aligned} \tag{55}$$

where the β terms that arise as the results of these integrals are defined as

$$\begin{aligned} \beta_{xy,1} &= C_{xy,1} + \frac{1}{\sigma_1}, \\ \beta_{xy,k} &= C_{xy,k} + \frac{1}{\sigma_k} + \frac{1}{\sigma_{k-1}} - \frac{1}{\sigma_{k-1}^2 \beta_{xy,k-1}}; \quad 1 < k < K \\ \beta_{xy,K} &= \frac{1}{\sigma_{K-1}} - \frac{1}{\sigma_{K-1}^2 \beta_{xy,K-1}}. \end{aligned} \tag{56}$$

If the particle identity changes, i.e. $n \neq m$, then the results of the integrals, β, are modified. This can be seen when calculating the integral

$$\begin{aligned} I_{m,2} I_{n,1} = \left(\prod_{n' \neq n \neq m=1}^{N} \int_{-\infty}^{\infty} \exp\left(-\beta^{(2)}_{xy,2} x_{n',2}^2 + \frac{2x_{n',3}}{\sigma_2} x_{n',2} - \frac{x_{n',3}^2}{\sigma_2} \right) dx_{n',2} \right) \\ \times \int_{-\infty}^{\infty} \exp\left(-\beta^{(1)}_{xy,2} x_{m,2}^2 - \frac{2x_{m,3}}{\sigma_2} x_{m,2} - \frac{x_{m,3}^2}{\sigma_2} \right) dx_{m,2} \int_{-\infty}^{\infty} \exp\left(-\beta_{xy,2} x_{n,2}^2 - \frac{2x_{n,3}}{\sigma_2} x_{n,2} - \frac{x_{n,3}^2}{\sigma_2} \right) dx_{n,2} \end{aligned} \tag{57}$$

which once integrated becomes

$$I_{m,2}I_{n,1} = \sqrt{\frac{\pi}{\beta_{xy,2}}}\sqrt{\frac{\pi}{\beta^{(1)}_{xy,2}}}\left(\frac{\pi}{\beta^{(2)}_{xy,2}}\right)^{\frac{N-2}{2}}\left(\prod_{n'\neq n\neq m=1}^{N}\exp\left(-\left(\frac{1}{\sigma_2}-\frac{1}{\sigma_2^2\beta^{(2)}_{xy,2}}\right)x^2_{n',3}\right)\right)\exp\left(-\left(\frac{1}{\sigma_2}-\frac{1}{\sigma_2^2\beta^{(1)}_{xy,2}}\right)x^2_{n,3}\right)$$
$$\times\exp\left(-\left(\frac{1}{\sigma_2}-\frac{1}{\sigma_2^2\beta_{xy,2}}\right)x^2_{n,3}\right) \quad (58)$$

When integrals from one time step to another do not match we pick up primed terms in β, such that the number of changes in particle identity, p, are tracked by the superscript $^{(p)}$, such that $\beta' = \beta^{(1)}$, $\beta'' = \beta^{(2)}$, etc. Additionally, even when these integrals are calculated in different orders we pick up the same numbers of primed and un-primed β terms. The primed terms are defined as

$$\begin{aligned}
\beta^{(1)}_{xy,1} &= C_{xy,1} + \frac{1}{\sigma_1},\\
\beta^{(p)}_{xy,k} &= C_{xy,k} + \frac{1}{\sigma_k} + \frac{1}{\sigma_{k-1}} - \frac{1}{\sigma^2_{k-1}\beta^{(p-1)}_{xy,k-1}}; \quad 1 < k < K,\\
\beta^{(p)}_{xy,K} &= \frac{1}{\sigma_{K-1}} - \frac{1}{\sigma^2_{K-1}\beta^{(p-1)}_{xy,K-1}},
\end{aligned} \quad (59)$$

with the $\beta_{z,k}$ and $\beta^{(p)}_{z,k}$ terms arising from the integrals in the z direction defined similarly. It can be shown that by using binomial expansions for the $\frac{1}{\sigma_{k-1}\beta_{xy,k-1}}$ terms that the β terms at each time step become

$$\beta^{(p)}_{xy,k} = \frac{1}{\sigma_k}\left(1 + a_{xy}\sigma_k\left((k-p) - \sum_{i=1}^{k} M_i\right)\right). \quad (60)$$

Finally, upon multiplication, summation, and removal of terms higher than second order in Δt_k, i.e. $(\Delta t_k)^3 \approx 0$, we get

$$P^{(x)}_{N,K} = \beta_{xy} \quad (61)$$

where

$$\beta_d = 1 - \frac{1}{2}a_{xy}\left(\sum_{k'=1}^{K-1} k'\sigma_{k'}\right) + \frac{1}{2}a_d N\left(\sum_{k'=1}^{K-1}\sigma_{k'}S_{k'}\right) - \frac{1}{2}a_d^2\left(\sum_{i,j=1}^{K-1}\sigma_i\sigma_j\right) \quad (62)$$

for $d = xy, z$, and $S_k = \sum_{k'=1}^{k}\mu_{k'}\Delta t_{k'}$. Therefore, the final result for the likelihood is

$$P(\boldsymbol{\Delta t}|\theta) \approx \frac{\beta^2_{xy}\beta_z}{V_{\text{eff}}}\left(\prod_{k=1}^{K-1}\mu_m N^3 e^{-\mu_m N\Delta t_k}\right) \quad (63)$$

Maximum approximate likelihood estimation

Maximum likelihood estimation procedure

The estimators, $\hat{\theta}^{(i)}_{\text{MAL}}$, are defined such that they maximise the approximate log-likelihood. Here we calculate this by finding the value of $\hat{\theta}^{(i)}_{\text{MAL}}$ that defines where the derivative with respect to each estimator is zero, i.e. $\partial_{\theta_i}\ln(P(\boldsymbol{\Delta t}|\theta)) = 0$. For ease we work with the log of the likelihood

$$\begin{aligned}\ln\left(P(\boldsymbol{\Delta t}|\theta)\right)) &= \ln\left(\mu_m^{K-1}\right) - \ln\left(V_{\text{box}}\right) - \mu_m N\left(\sum_{k=1}^{K-1}\Delta t_k\right) + \ln\left(P_{N,K}^{(x)}\right) + \ln\left(P_{N,K}^{(y)}\right)\ln\left(P_{N,K}^{(z)}\right) \\ &= \ln\left(\mu_m^{K-1}\right) - \ln\left(V_{\text{box}}\right) - \mu_m N\left(\sum_{k=1}^{K-1}\Delta t_k\right) + \ln\left(N^k\beta_{xy}\right) + \ln\left(N^k\beta_{xy}\right)\ln\left(N^k\beta_z\right) \\ &= \ln\left(\mu_m^{K-1}\right) - \ln\left(V_{\text{box}}\right) - \mu_m N\left(\sum_{k=1}^{K-1}\Delta t_k\right) + 3(K-1)\ln(N) + 2\ln(\beta_{xy}) + \ln(\beta_z)\end{aligned} \tag{64}$$

$$\partial_N \ln(\beta_{xy}) = \frac{a_{xy}\left(\sum_{k'=1}^{K-1}\sigma_{k'}S_{k'}\right)}{2\beta_{xy}} \tag{65}$$

Concentration

The estimator for the Molar concentration $\hat{C}_{\text{MAL}}$ is found from the estimator for the number of molecules, $\hat{N}_{\text{MAL}}$, within the effective confocal volume V_{eff} using the maximum approximate likelihood estimation procedure outlined previously. Specifically, partial differentiation of Equation 64 gives

$$\begin{aligned}0 &= \partial_N \ln\left(P(\boldsymbol{\Delta t}|\theta)\right) \\ &= -\mu_m\left(\sum_{k=1}^{K-1}\Delta t_k\right) + \frac{3(K-1)}{N} + \frac{a_{xy}\left(\sum_{k'=1}^{K-1}\sigma_{k'}S_{k'}\right)}{\beta_{xy}} + \frac{a_{xy}\left(\sum_{k'=1}^{K-1}\sigma_{k'}S_{k'}\right)}{2\beta_z} \\ &\approx -\mu_m\left(\sum_{k=1}^{K-1}\Delta t_k\right) + \frac{3(K-1)}{N}.\end{aligned} \tag{66}$$

where the last two terms have been neglected as β_{xy} and β_z are approximately 1 and the numerators are small for typical values of a_{xy}, a_z, and σ. Thus, the estimator for the molecular number within the confocal volume becomes

$$\begin{aligned}N &= \frac{3(K-1)}{\mu_m\left(\sum_{k=1}^{K-1}\Delta t_k\right)} \\ &= \frac{3}{\mu_m\langle\boldsymbol{\Delta t}\rangle},\end{aligned} \tag{67}$$

which may be converted into concentration via the effective volume V_{eff}, volume correction factor ε and Avrogadro's number as

$$\hat{C}_{\text{MAL}} = \frac{\hat{N}_{\text{MAL}}}{N_A V_{\text{eff}}} = \frac{\varepsilon}{N_A V_{\text{eff}}\mu_{\text{m}}\langle\boldsymbol{\Delta t}\rangle}. \tag{68}$$

The factor, ε, corrects the effective volume V_{eff} and must be determined for each experiment.

Diffusion

From simulations of synthetic FCS data, the estimator for the diffusion coefficient is close to

$$\hat{D}_{\text{MAL}} = \frac{\alpha(K-1)}{4(w_{xy}^{-2} + 2w_z^{-2})\mu_m N\sum_{i=0}^{K-1}\text{Cov}(\boldsymbol{\Delta t}_k, \boldsymbol{\Delta t}_{k+i})} + \eta, \tag{69}$$

where α and η are constants to be determined during calibration. When applying the same maximum likelihood procedure as before an estimator without these correction constants cannot currently be found.

4.3 Limitations of current FCS analyses

4.3.1 Low data density and high uncertainty

Fluorescence correlation spectroscopy (FCS) is used to determine the absolute number of fluorescent molecules in solution or cells by exploiting the information in the fluctuations of the number of photons observed in a very small confocal volume of ~ 1 fl, see Figure 4.1A. The intensity at time t, $I(t)$, is found from binning these collected photons, which can then be auto-correlated over successive lag times (τ) via the equation

$$G(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle} - 1. \tag{4.1}$$

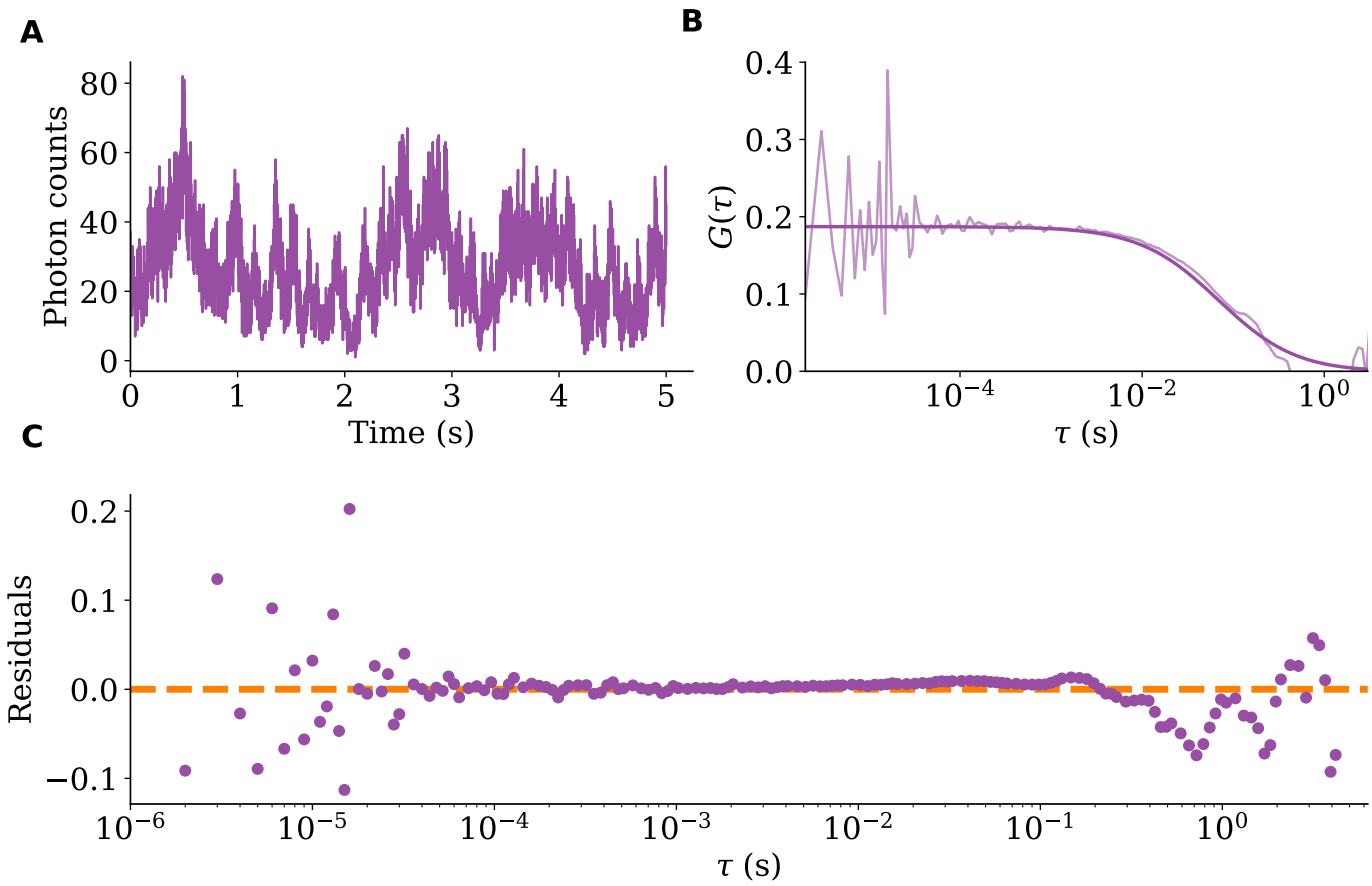

Figure 4.1: (**A**) Simulated photon counts binned into equal bins of 1 ms over a total of 5 s for a total of 1000 particles diffusing at $0.1\mu m^2 s^{-1}$. (**B**) Correlation curve generated from the count data in (**A**). (**C**) Residuals for the fit to data in (**B**) demonstrates poor fitting at short ($\tau < 10^{-4}$ s) and long ($\tau > 10^{-1}$ s) lag times.

This correlation is then fit to a model function to infer diffusion, concentration, and in some cases photo-kinetic parameters. Figure 4.1B shows a representative correlation curve and it's fit via a normal diffusion model

$$G(\tau) = \frac{1}{\langle N\rangle}\left(1+\frac{\tau}{\tau_D}\right)^{-1}\left(1+\left(\frac{w_{xy}}{w_z}\right)^2\frac{\tau}{\tau_D}\right)^{-\frac{1}{2}}, \tag{4.2}$$

for an average of N molecules, diffusion rate D, and confocal radii w_{xy} and w_z. Direct convolution of the full set of correlation values in Equation 4.1 is a computationally expensive task of order $\mathcal{O}(n^2)$ for n photons, and thus grows rapidly for increasing numbers photons. In practical applications of FCS, *multi-tau* algorithms are used to reduce down the number of calculations by binning the intensity data with progressively larger bin-widths following logarithmic density. The lag times start at the smallest possible, $\tau_0 \sim 1\mu s$, growing as powers of 2 to preserve detail at low lag times and improve

computational speed at longer lag times. Hence, multi-tau lowers memory usage and is much quicker, with a computational complexity of $\mathcal{O}(n \log(n))$, producing around a few hundred correlation points from the roughly $10^5 - 10^6$ photons collected in a typical FCS experiment (Schätzel et al., 1988). As FCS analysis infers parameters by fitting models to correlation points, the data density available for fitting is reduced by at least 3 orders of magnitude when using multi-tau algorithms. Thus, the reliability and accuracy of inferences is reduced as can be seen in the far left and right of the fit in Figure 4.1. These fitting issues manifest themselves in a number of ways, with the most pernicious being rogue data points at short and long lag times drastically altering the fit. Introducing arbitrary cut offs at both ends is not a viable way of rectifying this issue as any aberrations in remaining data points would make a larger contribution during fitting. Additionally, the necessary use of non-linear least squares regression to fit the data underestimates uncertainty on inferred parameters (Motulsky and Ransnas, 1987). Additional uncertainty in the data points is also introduced from using multi-tau as it averages over the data by re-binning it into increasingly larger bins (Saffarian and Elson, 2003). This manifests as shot noise at short lag times, see Figure 4.1C, and at longer lag times particles have moved out of the volume and thus the signal becomes uncorrelated. In summary, as correlation analyses explicitly depend on examining data over both small and large averages, FCS analysis is susceptible to poor inferences arising low data density and large uncertainties in some data points.

4.3.2 Pragmatic solutions

Previously, when fitting the FCS data for the DNA binding paper in chapter 2 and FCCS interaction measurements in chapter 3, I overcame these issues by increasing the number replicate FCS experiments so that poor fits may be discarded, as well as improving the reliability of fitting autocorrelation curves. I achieved this by implementing two rounds of fitting, an initial round using a genetic algorithm global optimiser (SciPy *differential evolution*) and a second round of gradient descent (SciPy *curve fit*), with the standard deviation on the data assigned according to the theoretical variance calculated by (Saffarian and Elson, 2003).

4.4 An alternative FCS framework

4.4.1 Probabilistic models of an FCS experiment

Distribution of photon counts

Jazani et al. (2019) defined that the number of observed photons within the time interval δt_k ($k = 1, 2, ..., K$) is drawn from a Poisson distribution

$$\gamma_k \sim \textbf{Poisson}(\lambda_k), \tag{4.3}$$

with mean λ_k. The parameter λ_k is a function of two components, the background photon emission rate, μ_b, and the sum of the molecular emissions, μ_k, as

$$\lambda_k = \delta t_k \left(\mu_b + \mu_k\right). \tag{4.4}$$

It should be noted that δt_k is the interval over which photons have been binned and can be variable to allow a minimum number of photons in each bin to be imposed. Unlike Jazani and colleagues, my analyses assume all time bins are of equal width and thus $\delta t = \delta t_k \ \forall \ k$. Molecular emissions, μ_k, is a sum of all molecules emitting detectable photons, with each molecule contributing to the signal dependent on its position with respect to the effective observation volume, otherwise known as the confocal volume (CV), according to the point spread function (PSF). By definition, the origin, i.e.

$(x, y, z) = (0, 0, 0)$ is located at the centre of the confocal volume. Thus, the molecular emissions may be approximated as

$$\mu_k = \mu_m \sum_{n=1}^{N} \mathrm{PSF}(x_{k,n}, y_{k,n}, z_{k,n}). \tag{4.5}$$

where N is the total number of molecules in the whole system. Here, μ_m defines a per molecule emission rate and encodes the likelihood of each particle emitting a photon per second and is thus dependent on the intensity of laser illumination (Qian, 1990). When measuring in a cell or solution in an FCS experiment N is a very large number of order of 10^{15} for a concentration of a few nM. Thankfully, the PSF is a sharply peaked function, thus only molecules close to the centre of the PSF contribute to this summation. Jazani et al. (2019) considered a finite number of molecules which may contribute to the signal, N_{Sim}, selected by a binary value on each simulated molecule, b_n. The molecular number can then be inferred from the number of b_n values equal to 1. In summary, the model for the mean photons emitted in the interval k in Equation 4.4 becomes

$$\mu_k = \delta t_k \left(\mu_b + \mu_m \sum_{n=1}^{N_{\mathrm{Sim}}} b_n \mathrm{PSF}(x_{k,n}, y_{k,n}, z_{k,n}) \right), \tag{4.6}$$

Typically, the PSF is defined as a 3D Gaussian confocal observation volume with the signal contribution of each molecule as a function of position given by

$$\mathrm{PSF}(x, y, z) = \exp\left(-2 \left(\frac{x^2}{w_{xy}^2} + \frac{y^2}{w_{xy}^2} + \frac{z^2}{w_z^2} \right) \right) \tag{4.7}$$

where w_{xy} and w_z denote the characteristic radii in the x-y and z directions of the observation volume respectively whereby probability of detecting a photon emitted at that position drops to e^{-2} when the displacement of the particle is equal to these radii. Clearly, the PSF does not have a hard cut off and hence the effective volume

$$V_{\mathrm{eff}} = \pi^{\frac{3}{2}} C w_{xy}^2 w_z \tag{4.8}$$

is used to calculate concentrations. Note that a correction factor C is dependent on the optics has been added as the effective volume is typically larger than $\pi^{\frac{3}{2}} w_{xy}^2 w_z$. The correction factor, C, may be found for each FCS setup by using a known concentration of fluorescent molecules in solution to calibrate V_{eff}.

Time between photon arrivals

Tavakoli et al. (2020), proposed an updated model to Jazani et al. (2019) which infers the same molecular parameters but from the individual photon arrival times. The arrival times are the rawest form of data from the detectors in an FCS experiment. These data can be used to create an array times between arrivals without arbitrary binning and pre-processing as

$$\underline{\Delta t} = (\Delta t_1, \Delta t_2, ..., \Delta t_{K-1}) \tag{4.9}$$

where $\Delta t_k = t_{k+1} - t_k$ for a total of K photons collected. The times between successive photons is the waiting period between two Poisson distributed events and as such is drawn from an exponential distribution

$$\Delta t_k \sim \mathbf{Exponential}(\mu_k), \tag{4.10}$$

with μ_k having the same interpretation and form as in the Jazani model (Equation 4.6) except for

the lack of δt, namely

$$\mu_k = \mu_b + \mu_m \sum_{n=1}^{N} \mathrm{PSF}(x_{k,n}, y_{k,n}, z_{k,n}). \tag{4.11}$$

Diffusion

In 1 dimension the diffusion of a particle labelled with position x at times indexed by k is governed by Fick's second law (Fick, 1855; Du and Kou, 2020)

$$\frac{\partial f(x,t)}{\partial t} = D \frac{\partial^2 f(x,t)}{\partial x^2}, \tag{4.12}$$

where $f(x,t)$ is the probability density of finding a particle at position x at time t. When considering no spatial boundaries and starting at the origin, a solution to Equation 4.12 is

$$f(x,t) = \frac{1}{\sqrt{4\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right). \tag{4.13}$$

The particle has an equal probability of jumping left or right and as such the mean displacement from the origin, $\langle x(t) \rangle$, is zero. Only, the variance of this distribution, also know as the mean squared displacement (MSD) increases over time, defined by

$$\mathrm{MSD} = \langle x(t)^2 \rangle = \int_{-\infty}^{\infty} x^2 f(x,t) dx = 2Dt. \tag{4.14}$$

When considering more complex dynamics than normal Brownian motion, the MSD will depend on time in a non-linear fashion, with examples being confinement, interactions, and active transport. The probability density of the particles position at a later time $t_{k'}$ is

$$f(x_k, t_{k'}) = \frac{1}{\sqrt{4\pi D(t_{k'} - t_k)}} \exp\left(-\frac{(x_{k'} - x_k)^2}{4D(t_{k'} - t_k)}\right). \tag{4.15}$$

Thus, the position of a particle at a later time $t_{k'}$ representing a time step of $\Delta t_k = t_{k'} - t_k$ is normally distributed with mean x_k and variance $2D\Delta t_k$. Which may be rewritten as

$$f(x_k | x_{k'}, D) = \mathbf{Normal}(x_{k'}, 2\Delta t_k D). \tag{4.16}$$

This is easily extended to three dimensions as each direction can be considered to be independent and as such is drawn from its own independent and identical normal distribution, forming a multivariate normal distribution with zero covariance between each direction.

Photon emission rates depend on concentration and diffusion

Photon emission rates drawn from a Poisson (Equation 4.3) conditioned on its mean (Equation 4.5) are a sum over the contribution of each molecule determined from the point spread function on their positions with respect to the centre of the confocal volume. Thus, only molecules close to the centre of illumination within the confocal volume will significantly contribute to the number of observed photons. Intuitively, the brighter the signal the more molecules must be present. However, dynamical information about the movement of the molecules, encoded by its diffusion coefficient, D, can only come from changes in the point spread function. Monte Carlo simulations simulations of FCS experiments demonstrate that the distribution of PSF values follows a Poisson distribution with mean and variance dependent on the number of molecules, Figure 4.2A. This is in agreement with Photon Counting Histogram (PCH) analysis (Müller et al., 2000) which proposed that the

distribution of molecules in a small volume situated within a larger reservoir is

$$N \sim \textbf{Poisson}(N_{CV}), \tag{4.17}$$

where N_{CV} is the average number of molecules what would be found in the observation volume. If the molecules in solution are homogeneous then N_{CV} will be equal to the average concentration of molecules in solution.

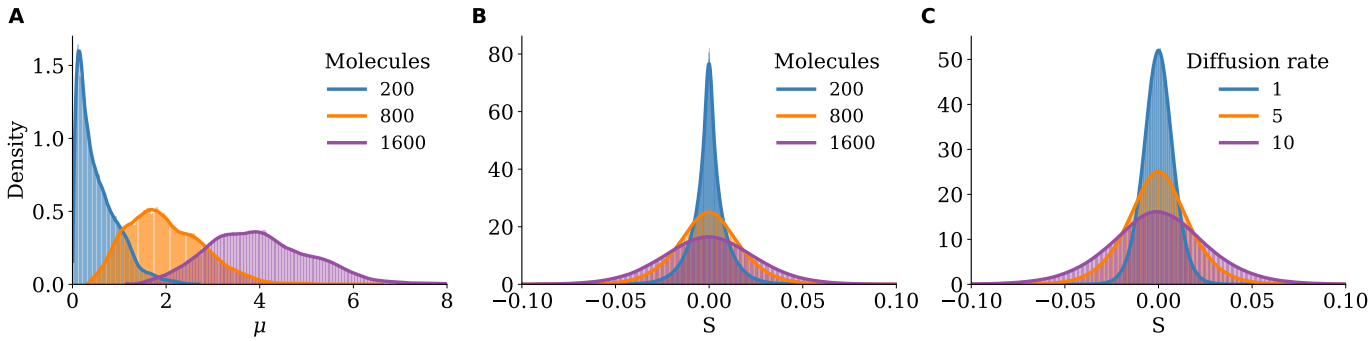

Figure 4.2: (**A**) Histograms of photon emission rates, μ, for simulations over three different molecular concentrations: $N_{Tot} = 200$, $N_{Tot} = 800$, and $N_{Tot} = 1600$. These closely follow a Poisson distribution with means dependent on molecular abundance. (**B**) The distribution of S for increasing numbers of total molecules and a lag of 1 bin width, δt. Variance increases for higher concentrations of molecules within the confocal Volume. (**C**) The variance of S also increases with diffusion rates ranging from 1 $\mu m^2 s^{-1}$ to 10 $\mu m^2 s^{-1}$. Time lag is again equal to δt.

For any snap shot in time the positions of the molecules can be considered to be drawn from a uniform distribution in each dimension. Instinctively, diffusion should cause changes to the value of each PSF in time and in turn the photon emission rates such that time shifted sums of PSFs over the molecules within the volume will vary over time. Defining

$$S_k = \frac{\mu_{k+1} - \mu_k}{\mu_{\mathrm{m}}} \quad k = 1, 2, ..., K-1 \tag{4.18}$$

as the shifted contribution to the signal, the effect of different molecular numbers of total simulated molecules can be investigated. The distribution of S is approximately normal, with increasing variance from higher numbers of molecules (Figure 4.2B) and faster diffusion rates (Figure 4.2C). The change between μ_k and μ_{k+1} is due to changes in the magnitude of the PSF at each of these times and is the brought about by molecular movement. It should be noted that the PSF is symmetric in at least 2 axes, usually x-y. Hence, there are some symmetries and configurations of molecular numbers and diffusion rates that appear the same. Namely, slower moving molecules around the periphery of the observable volume will appear similar to a small number of fast moving molecules in the centre. Monte Carlo simulations of the variance of S suggests that diffusion rates and molecular numbers may be approximately calculated from the variance of S as

$$\mathrm{Var(S)} \propto D N_{\mathrm{Tot}} \Delta t \tag{4.19}$$

where Var(S) is the variance of the distribution of S, N_{Tot} is the total number of molecules simulated, and $\Delta t = \delta t(k' - k)$. Variance in this shift Var(S) is strongly linear with N_{Tot} and D following a gradient of ≈ 0.088, Figure 4.3. The gradient value likely corresponds to missing constants to be found and to the ratio of the effective observation volume V_{eff} to bounding volume used during simulations, which relates the actual number of particles in the volume N_{CV} to total simulated N_{Tot}. In reality we do not have direct access to the PSF value over time and thus to reliably infer molecular numbers as well as diffusion rates, a formal model is required that can equate shifts in the PSF to the observed photon count data.

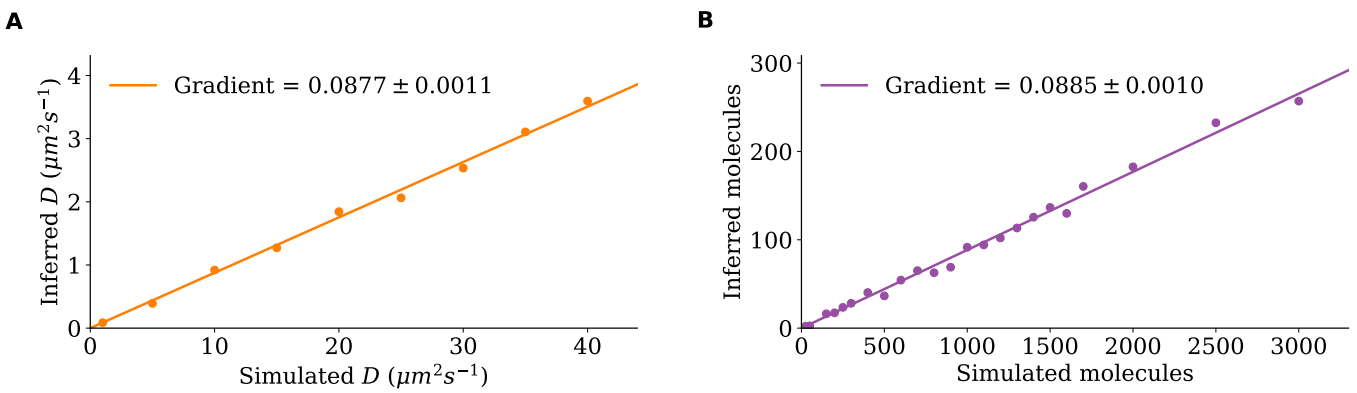

Figure 4.3: (**A**) Inferred diffusion rates D and molecular numbers (**B**) calculated via Equation 4.19 shows a linear relationship. $N_{\text{Tot}} = 800$ for simulations in (**A**) and $D = 5\mu\text{m}^2\text{s}^{-1}$ in (**B**).

4.4.2 The properties of changes in photon counts

Pulling this knowledge together, the changes in photon counts can be related to the physical parameters of the system. Photons binned in intervals of δt labelled with index k are Poisson distributed as $\gamma_k \sim \textbf{Poisson}(\mu_k)$. To observe dynamic properties, the differences in photon counts between k and k' over a time $\delta t(k' - k)$ is denoted as

$$\Delta_{k,k'} = \gamma_{k'} - \gamma_k. \tag{4.20}$$

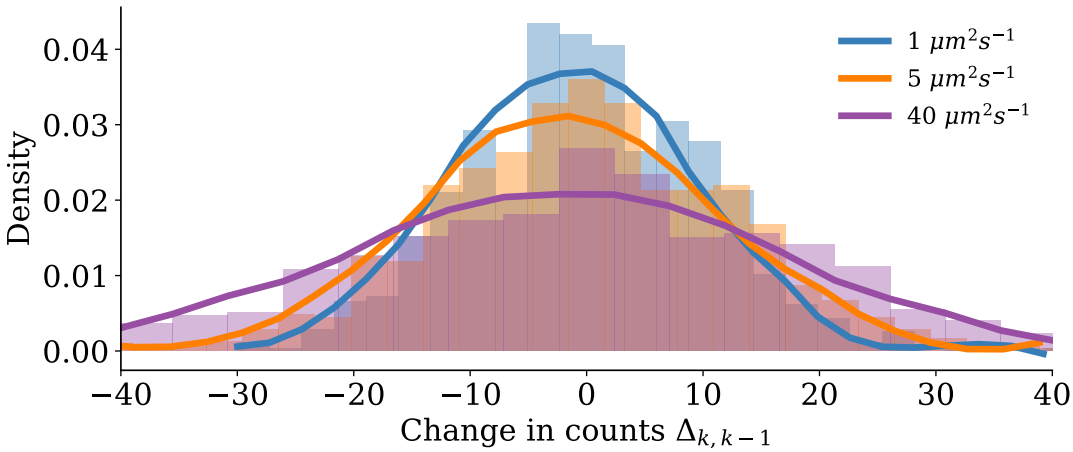

Figure 4.4: Distribution of differences photon counts in interval $\delta t = 1$ ms from simulations with $N_{\text{Total}} = 800$ over three diffusion rates: 1 $\mu\text{m}^2\text{s}^{-1}$, 5 $\mu\text{m}^2\text{s}^{-1}$, and 40 $\mu\text{m}^2\text{s}^{-1}$. The change in photon counts has a mean of zero and a variance that increases with diffusion rate.

When calculated, this shift in photon counts is a difference between two Poisson distributions, otherwise known as the Skellam distribution (Skellam, 1946) with mean $\mu_{k'} - \mu_k$ and variance $\mu_{k'} + \mu_k$. This distribution maps well to simulations of the change in counts as seen in Figure 4.4. To confirm this is correct I have calculated the mean and variance of $\Delta_{k,k'}$ in terms of the physical parameters; molecular number N, diffusion coefficient D, molecular brightness μ_m, and background emission rate μ_b. As before, the mean of the photons within each bin is μ_k, and for brevity I define $\Gamma_k = \text{PSF}(x_{n,k}, y_{n,k}, z_{n,k})$.

Mean change in photon counts

The expectation value of $\Delta_{k,k'}$ is found to be

$$\text{E}[\Delta_{k,k'}] = \text{E}[\gamma_{k'} - \gamma_k] = E[\gamma_{k'}|\mu_{k'}] - \text{E}[\gamma_k|\mu_k] = E[\mu_{k'}] - \text{E}[\mu_k] = 0, \tag{4.21}$$

using the Law of Iterated Expectations (LIE) (Billingsley, 2008), $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]]$, with

$$\mathrm{E}[\gamma_k] = \mathrm{E}[\mathrm{E}[\gamma_k|\mu_k]] = \mathrm{E}[\mathrm{E}[\mathbf{Poi}(\mu_k)|\mu_k]] = \mathrm{E}[\mu_k] \tag{4.22}$$

because the mean of a Poisson distribution is simply its parameter μ_k. As it is assumed that no changes in the molecular emission parameters, diffusion rates or average number of molecules occur over the experiment, it follows that the mean of μ_k is is same over all k. Hence, $E[\Delta_{k,k'}] = 0$ is to be expected because on average there should the same number of counts from one interval to another and therefore the average of the difference should be zero up to some amount of drift. This can be seen in Figure 4.5A.

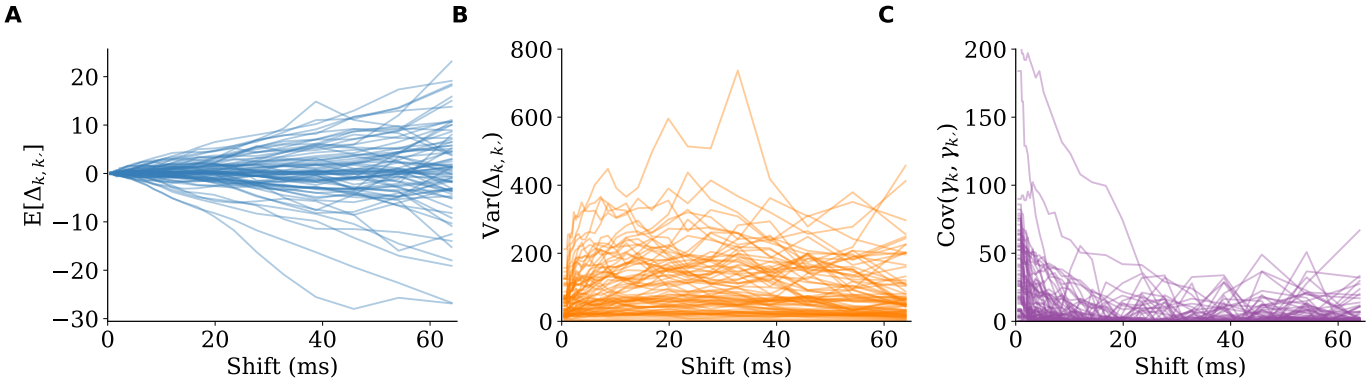

Figure 4.5: The mean (**A**) and variance (**B**) of $\Delta_{k,k'} = \gamma_{k'} - \gamma_k$ over increasing values of the shift $\delta t(k' - k)$ from Monte Carlo simulations of FCS experiments over a range of diffusion rates from $D = 0.1 - 10$ µm^{2}s^{-1} and total molecules $N_{\mathrm{Tot}} = 500 - 4000$. The mean is on average zero with some drift seen in individual FCS experiments. $\mathrm{Var}(\Delta_{k,k'})$ is zero for a zero time shift and grows quickly for larger time shifts. (**C**) The covariance between γ_k and $\gamma_{k'}$ starts high and rapidly decreases as time between k and k' increases.

This may be verified by computing the expectation using LIE and the definition of μ_k given in Equation 4.5 to yield

$$\begin{aligned}
\mathrm{E}[\mu_k] &= \mathrm{E}[\mathrm{E}[\mu_k|\mu_b, \mu_m, N, \Gamma_k]] \\
&= \mathrm{E}\left[\delta t \left(\mu_b + \mu_m \sum^{N} \Gamma_k\right)\right] \\
&= \delta t \left(\mathrm{E}[\mu_b] + \mathrm{E}[\mu_m]\mathrm{E}\left[\sum^{N} \Gamma_k\right]\right) \\
&= \delta t \left(\mathrm{E}[\mu_b] + \mathrm{E}[\mu_m]\mathrm{E}[N]\mathrm{E}\left[\Gamma_k\right]\right).
\end{aligned} \tag{4.23}$$

Here it is assumed that the mean sum over Γ_k is simply the expected number of molecules N multiplied by $\mathrm{E}[\Gamma]$. It is reasonable to assume that the distributions for the number of molecules, background rate, molecular brightness and molecular positions (encoded in Γ) are all independent of one another and are identically distributed random variables (i.i.d) over all intervals k during the experiment. Additionally, in the case of a Gaussian point spread function then $\mathrm{E}[\Gamma_k]$ becomes

$$\begin{aligned}
\mathrm{E}[\Gamma_k] &= \iiint_V \exp\left(-\frac{2x_k^2}{w_{xy}^2}\right) \exp\left(-\frac{2y_k^2}{w_{xy}^2}\right) \exp\left(-\frac{2z_k^2}{w_z^2}\right) \frac{1}{L_x}\frac{1}{L_y}\frac{1}{L_z} dxdydz \\
&\simeq \frac{1}{V} \iiint_{-\infty}^{\infty} \exp\left(-\frac{2x_k^2}{w_{xy}^2}\right) \exp\left(-\frac{2y_k^2}{w_{xy}^2}\right) \exp\left(-\frac{2z_k^2}{w_z^2}\right) dxdydz \\
&\simeq \frac{\pi^{\frac{3}{2}} w_{xy}^2 w_z}{2^{\frac{3}{2}} V}
\end{aligned} \tag{4.24}$$

where I have made use of the Law of the Unconscious Statistician (LOTUS) in the first line, with the positions in each direction distributed according to a uniform distribution over the bounding box,

i.e. $x_k \sim \mathbf{Uniform}(-\frac{L_x}{2}, \frac{L_x}{2})$. Despite total volume in an FCS experiment being finite and defined as $V = L_x L_y L_z$, the definite integral may be approximated and expanded over space to infinity as the contribution of the Gaussian terms are negligible outside the confocal volume provided $L_{xy} \gg w_{xy}$ and $L_z \gg w_z$. According to my Monte Carlo simulations, this approximation holds for $\frac{L_i}{w_i} > 4$, below which edge effects are observed. In the case of a solution or large cellular compartment such as the nucleus, the volume to effective volume V_{eff} is much greater than 4 and as such the approximation holds up well. We may therefore conclude that $\mathrm{E}[\Gamma_k]$ is a constant and independent of the interval k such that

$$\mathrm{E}[\Gamma_k] = \mathrm{E}[\Gamma_{k'}] \ \forall \ k. \tag{4.25}$$

Hence, from this and Equation 4.23 it can be concluded,

$$\mathrm{E}[\mu_k] = \mathrm{E}[\mu_{k'}] \ \forall \ k. \tag{4.26}$$

Variance of change in photon counts

As previously discussed, in contrast to the mean, the variance of $\Delta_{k,k'}$ is non-zero due to particles diffusing within in the confocal volume. This can be seen in Figure 4.5B-C where variance of $\Delta_{k,k'}$ and covariance between the photon counts γ_k and $\gamma_{k'}$ are non-zero and are functions of the time shift $\delta(k'-k)$. Using the Law of Iterated Variances (Weiss, 2005) $\mathrm{Var}(Y) = \mathrm{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathrm{E}[Y|X])$,

$$\mathrm{Var}(\Delta_{k,k'}) = \mathrm{E}[\mathrm{Var}(\Delta_{k,k'}|\mu_k, \mu_{k'})] + \mathrm{Var}(\mathrm{E}[\Delta_{k,k'}|\mu_k, \mu_{k'}]), \tag{4.27}$$

with the first term expanded as

$$\begin{aligned}\mathrm{E}[\mathrm{Var}(\Delta_{k,k'}|\mu_k, \mu_{k'})] =& \mathrm{E}[\mathrm{Var}(\gamma_k|\mu_k, \mu_{k'}) + \mathrm{Var}(\gamma_{k'}|\mu_k, \mu_{k'}) \\ &- 2\mathrm{Cov}(\gamma_k, \gamma_{k'}|\mu_k, \mu_{k'})] \\ =& \mathrm{E}[\mu_k + \mu_{k'} - 2\mathrm{Cov}(\gamma_k, \gamma_{k'}|\mu_k, \mu_{k'})],\end{aligned} \tag{4.28}$$

making use of the property $\mathrm{Var}(\gamma|\mu) = \mathrm{Var}(\mathbf{Poi}(\mu)) = \mu$. To complete this part of the calculation, the covariance between the counts is

$$\begin{aligned}\mathrm{Cov}(\gamma_k, \gamma_{k'}|\mu_k, \mu_{k'}) &= \mathrm{E}[\gamma_k\gamma_{k'}|\mu_k, \mu_{k'}] - \mathrm{E}[\gamma_k|\mu_k]\mathrm{E}[\gamma_k|\mu_{k'}] \\ &= \mathrm{E}[\gamma_k(\gamma_{k'} - \gamma_k + \gamma_k)|\mu_k, \mu_{k'}] - \mu_k\mu_{k'} \\ &= \mathrm{E}[\gamma_k(\gamma_{k'} - \gamma_k)|\mu_k, \mu_{k'}] + \mathrm{E}[\gamma_k^2|\mu_k] - \mu_k\mu_{k'} \\ &= \mathrm{E}[\gamma_k|\mu_k]\mathrm{E}[(\gamma_{k'} - \gamma_k)|\mu_k, \mu_{k'}] + \mu_k^2 + \mu_k - \mu_k\mu_{k'} \\ &= \mu_k(\mu_{k'} - \mu_k) + \mu_k^2 + \mu_k - \mu_k\mu_{k'} \\ &= \mu_k\end{aligned} \tag{4.29}$$

where I have made use of the fact that the difference of two Poisson distributed processes is independent of the photon counts. Hence, Equation 4.28 becomes

$$\begin{aligned}\mathrm{E}[\mathrm{Var}(\Delta_{k,k'}|\mu_k, \mu_{k'})] &= \mathrm{E}[\mu_k + \mu_{k'} - 2\mathrm{Cov}(\gamma_k, \gamma_{k'}|\mu_k, \mu_{k'})] \\ &= \mathrm{E}[\mu_k + \mu_{k'} - 2\mu_k] \\ &= \mathrm{E}[\mu_{k'} - \mu_k] \\ &= \mathrm{E}[\mu_{k'}] - \mathrm{E}[\mu_k] \\ &= 0\end{aligned} \tag{4.30}$$

This is again zero as the mean photon counts between intervals k and k' are the same. Unlike the

first term, the second term of the variance in Equation 4.27 is non-zero and can be expanded as

$$\begin{aligned}\mathrm{Var}(\mathrm{E}[\Delta_{k,k'}|\mu_k,\mu_{k'}]) =&\mathrm{Var}(\mu_k-\mu_{k'})\\ =&\mathrm{Var}(\mu_k)+\mathrm{Var}(\mu_{k'})-2\mathrm{Cov}(\mu_k,\mu_{k'})\\ =&\underbrace{\mathrm{E}[\mu_k^2]-\mathrm{E}[\mu_k]^2}_{A}+\underbrace{\mathrm{E}[\mu_{k'}^2]-\mathrm{E}[\mu_{k'}]^2}_{B}\\ &-2(\underbrace{\mathrm{E}[\mu_k\mu_{k'}]-\mathrm{E}[\mu_k]\mathrm{E}[\mu_{k'}]}_{C})\end{aligned} \tag{4.31}$$

where A is

$$\begin{aligned}A =&\mathrm{E}\left[(\delta t)^2\left(\mu_b+\mu_m\sum^N\Gamma_k\right)^2\right]-\mathrm{E}\left[\delta t\left(\mu_b+\mu_m\sum^N\Gamma_k\right)\right]^2\\ =&(\delta t)^2\left(\mathrm{E}\left[\left(\mu_b+\mu_m\sum^N\Gamma_k\right)^2\right]-\mathrm{E}\left[\left(\mu_b+\mu_m\sum^N\Gamma_k\right)\right]^2\right)\\ =&(\delta t)^2(\mathrm{E}[\mu_b^2+2\mu_b\mu_m N\Gamma_k+\mu_m^2N^2\Gamma_k^2]-\mathrm{E}[\mu_b]^2\\ &-2\mathrm{E}[\mu_b]\mathrm{E}[\mu_m N\Gamma_k]-\mathrm{E}[\mu_m N\Gamma_k]^2)\\ =&(\delta t)^2\left(\mathrm{Var}(\mu_b)+\mathrm{Var}(\mu_m N\Gamma_k)\right).\end{aligned} \tag{4.32}$$

Here, it is again assumed that the expectation of the sum is

$$\begin{aligned}\mathrm{E}\left[\sum^N\Gamma_k\right]\simeq&\mathrm{E}\left[\Gamma_k\sum^N 1\right]\\ \simeq&\mathrm{E}[N]\mathrm{E}[\Gamma_k]\end{aligned} \tag{4.33}$$

as Γ_k is not dependent on the sum variable, hence $\mathrm{E}[\Gamma_k]$ is simply multiplied by the expected number of molecules $\mathrm{E}[N]$. B is expanded similarly and by $\mathrm{E}[\Gamma_k]=\mathrm{E}[\Gamma_{k'}]$ it can be concluded that $A=B$. Again expanding in the same manner as A, the last term C is

$$\begin{aligned}C =&(\delta t)^2(\mathrm{E}[\mu_b^2+\mu_b\mu_m N\Gamma_k+\mu_b\mu_m N\Gamma_{k'}+\mu_m^2N^2\Gamma_k\Gamma_{k'}]\\ &-\mathrm{E}[\mu_b]^2-\mathrm{E}[\mu_b]\mathrm{E}[\mu_m N\Gamma_k]-\mathrm{E}[\mu_b]\mathrm{E}[\mu_m N\Gamma_{k'}]-\mathrm{E}[\mu_m N]^2\mathrm{E}[\Gamma_k]\mathrm{E}[\Gamma_{k'}])\\ =&(\delta t)^2(\mathrm{E}[\mu_b^2]-\mathrm{E}[\mu_b]^2+\mathrm{E}[\mu_m^2N^2]\mathrm{E}[\Gamma_k\Gamma_{k'}]-\mathrm{E}[\mu_m N]^2\mathrm{E}[\Gamma_k]\mathrm{E}[\Gamma_{k'}])\\ =&(\delta t)^2(\mathrm{Var}(\mu_b)+\mathrm{E}[\mu_m^2N^2]\mathrm{E}[\Gamma_k\Gamma_{k'}]-\mathrm{E}[\mu_m N]^2\mathrm{E}[\Gamma_k]\mathrm{E}[\Gamma_{k'}]).\end{aligned} \tag{4.34}$$

Higher moments of Γ_k, namely $\mathrm{E}[\Gamma_k^2]$, possess the same property as $\mathrm{E}[\Gamma_k]$, see Equation 4.25, and as such

$$\begin{aligned}\mathrm{Var}(\mathrm{E}[\Delta_{k,k'}|\mu_k,\mu_{k'}]) =&A+B-2C\\ =&2(\delta t)^2(\mathrm{Var}(\mu_m N\Gamma_k)-\mathrm{E}[\mu_m^2N^2]\mathrm{E}[\Gamma_k\Gamma_{k'}]\\ &+\mathrm{E}[\mu_m N]^2\mathrm{E}[\Gamma_k]\mathrm{E}[\Gamma_{k'}])\\ =&2(\delta t)^2(\mathrm{E}[\mu_m^2N^2\Gamma_k^2]-\mathrm{E}[\mu_m N\Gamma_k]^2\\ &-\mathrm{E}[\mu_m^2N^2\Gamma_k\Gamma_{k'}]+\mathrm{E}[\mu_m N\Gamma_k]^2)\\ =&2(\delta t)^2\mathrm{E}[\mu_m^2N^2](\mathrm{E}[\Gamma_k^2]-\mathrm{E}[\Gamma_k\Gamma_{k'}]).\end{aligned} \tag{4.35}$$

The term $\mathrm{E}[\mu_m^2N^2]$ is known and it can be readily calculated that $\mathrm{E}[\Gamma_k^2]=\frac{\pi^{\frac{3}{2}}w_{xy}^2w_z}{8V}$. However, to complete the calculation $\mathrm{E}[\Gamma_k\Gamma_{k'}]$ must be determined. Progress can be made by noting that each direction is independent of one another and the position at a later interval k' is random variable drawn as $x_{k'}=x_k+\mathbf{Normal}(0,\sigma^2)$ from the earlier position during the interval k. The y and z directions are similarly drawn from their respective previous positions with the same variance of σ

as the particle may move freely in all directions at the same rate. The variance of these normal distributions are related to the diffusion rate as $\sigma^2 = 2D\delta t(k'-k)$ and using $\mathcal{N}_i = \mathbf{Normal}(0,\sigma^2)$ to denote a normal distribution used to draw random variables in the direction i, the term $\mathrm{E}[\Gamma_k\Gamma_{k'}]$ becomes

$$
\begin{aligned}
&= \mathrm{E}\left[\iiint_V e^{-\frac{2(x_k^2+x_{k'}^2)}{w_{xy}^2}} e^{-\frac{2(y_k^2+y_{k'}^2)}{w_{xy}^2}} e^{-\frac{2(z_k^2+z_{k'}^2)}{w_z^2}} \frac{1}{L_xL_yL_z} dx'dy'dz'\right] \; * \\
&= \frac{1}{V}\mathrm{E}\left[e^{-\frac{2\mathcal{N}_x^2}{w_{xy}^2}} e^{-\frac{2\mathcal{N}_y^2}{w_{xy}^2}} e^{-\frac{2\mathcal{N}_z^2}{w_z^2}} \iiint_{-\infty}^{\infty} e^{-\frac{4(x_{k'}^2+\mathcal{N}_x x_{k'})}{w_{xy}^2}} e^{-\frac{4(y_{k'}^2+\mathcal{N}_y y_{k'})}{w_{xy}^2}} e^{-\frac{4(z_{k'}^2+\mathcal{N}_z z_{k'})}{w_z^2}} dxdydz\right] \\
&= \frac{1}{V}\mathrm{E}\left[e^{-\frac{2\mathcal{N}_x^2}{w_{xy}^2}} e^{-\frac{2\mathcal{N}_y^2}{w_{xy}^2}} e^{-\frac{2\mathcal{N}_z^2}{w_z^2}} \frac{\sqrt{\pi}w_{xy}}{2} e^{\frac{\mathcal{N}_x}{w_{xy}^2}} \frac{\sqrt{\pi}w_{xy}}{2} e^{\frac{\mathcal{N}_y}{w_{xy}^2}} \frac{\sqrt{\pi}w_z}{2} e^{\frac{\mathcal{N}_z}{w_z^2}}\right] \\
&= \frac{\pi^{\frac{3}{2}} w_{xy}^2 w_z}{8V}\mathrm{E}\left[e^{-\frac{\mathcal{N}_x^2}{w_{xy}^2}} e^{-\frac{\mathcal{N}_y^2}{w_{xy}^2}} e^{-\frac{\mathcal{N}_z^2}{w_z^2}}\right] \; \dagger \\
&= \frac{\pi^{\frac{3}{2}} w_{xy}^2 w_z}{8V} \int_{-\infty}^{\infty} e^{-\frac{x^2}{w_{xy}^2}} \cdot \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{w_{xy}^2}} \cdot \frac{e^{-\frac{y^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dy \int_{-\infty}^{\infty} e^{-\frac{z^2}{w_z^2}} \cdot \frac{e^{-\frac{z^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dz \\
&= \frac{\pi^{\frac{3}{2}} w_{xy}^2 w_z}{8V} \left(\frac{1}{\sqrt{1+\frac{2\sigma^2}{w_{wxy}^2}}}\right)\left(\frac{1}{\sqrt{1+\frac{2\sigma^2}{w_{wxy}^2}}}\right)\left(\frac{1}{\sqrt{1+\frac{2\sigma^2}{w_z^2}}}\right)
\end{aligned}
\tag{4.36}
$$

where on the first line ($*$) LOTUS was used for x_k, y_k, z_k, and on the fourth line ($\dagger$) LOTUS was applied to $\mathcal{N}_x, \mathcal{N}_y, \mathcal{N}_z$. Finally, by substituting in the definition for σ^2, the variance of the difference in binned photon counts, Equation 4.27, between intervals k and k' becomes

$$
\mathrm{Var}(\Delta_{k,k'}) = \frac{\pi^{\frac{3}{2}} w_{xy}^2 w_z (\delta t)^2}{4V}\mathrm{E}[\mu_m^2]\mathrm{E}[N^2]\left(1 - \frac{1}{1+\frac{4D\delta t\Delta k}{w_{xy}^2}}\frac{1}{\sqrt{1+\frac{4D\delta t\Delta k}{w_z^2}}}\right), \tag{4.37}
$$

for $\Delta k = k' - k$.

Variance in $\Delta_{k,k'}$ is simply related to Var(S)

$\mathrm{Var}(\Delta_{k,k'})$ may be approximated using a binomial expansion when $\frac{4D\delta t\Delta k}{w^2}$ is small as

$$
\begin{aligned}
\mathrm{Var}(\Delta_{k,k'}) &\approx \frac{\pi^{\frac{3}{2}} w_{xy}^2 w_z (\delta t)^2}{4V}\mathrm{E}[\mu_m^2]\mathrm{E}[N^2]\left(1 - \left(1 - \frac{4D\delta t\Delta k}{w_{xy}^2}\right)\left(1 - \frac{1}{2}\frac{4D\delta t\Delta k}{w_z^2}\right)\right) \\
&\approx \frac{\pi^{\frac{3}{2}} w_{xy}^2 w_z (\delta t)^2}{4V}\mathrm{E}[\mu_m^2]\mathrm{E}[N^2]\left(\frac{4D\delta t\Delta k}{w_{xy}^2} + \frac{1}{2}\frac{4D\delta t\Delta k}{w_z^2} + \mathrm{H.O.Ts}\right) \\
&\approx \frac{\pi^{\frac{3}{2}} w_{xy}^2 w_z (\delta t)^3 D\Delta k}{V}\mathrm{E}[\mu_m^2]\mathrm{E}[N^2]\left(\frac{1}{w_{xy}^2} + \frac{1}{2w_z^2}\right)
\end{aligned}
\tag{4.38}
$$

where higher order terms (H.O.Ts) that are quadratic in $\left(\frac{4D\delta t\Delta k}{w^2}\right)^2$ have been neglected as they are very small. The photon counts are related to the point spread function by $\delta t\mu_m$ and as such to obtain a theoretical approximation of Var(S) then Equation 4.38 should be divided by $(\delta t)^2\mathrm{E}[\mu_m]^2$ to get

$$
\mathrm{Var(S)} \approx \frac{\pi^{\frac{3}{2}} w_{xy}^2 w_z N D\Delta t}{V}\left(\frac{1}{w_{xy}^2} + \frac{1}{2w_z^2}\right), \tag{4.39}
$$

as $\mathrm{Var}(bX) = b^2\mathrm{Var}(X)$ for a constant b. For the particular parameters used during my Monte Carlo

simulations, $V = 96$ µm^3, $w_{xy} = 0.3$µm, and $w_z = 1.5$µm then Var(S) is approximately

$$\mathrm{Var(S)} \approx 0.0887 DN\Delta t. \tag{4.40}$$

This has the same linear relationship on N and D as the approximation made before in Equation 4.19 and has a constant of proportionality that is close to those found from simulations (0.0877 and 0.0885), see Figure 4.3. In summary, we know that the change in photon counts counts follows a Skellam distribution, with a mean of zero and a variance that scales with molecular numbers, N, and diffusion rate, D, see Equation 4.40. Furthermore, as the time interval between observed counts increases, they become less correlated and their differences vary more broadly in a linear fashion (Figure 4.6).

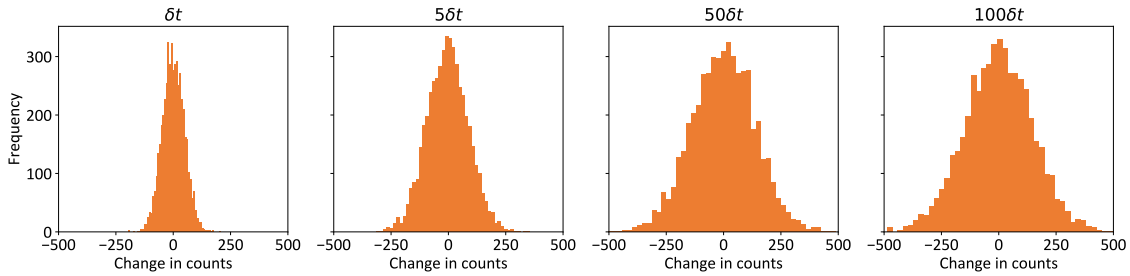

Figure 4.6: Skellam distribution with mean zero and variance as calculated from Equation 4.37 for the parameters $N_{\mathrm{tot}} = 500, D = 5$ µm^2/s, $\mu_{\mathrm{mol}} = 22500$ s^{-1}, $w_{xy} = 0.3$ µm, $w_z = 1.5$ µm, $L_{xy} = 4$ µm, $L_z = 6$ µm, $\delta t = 50$ µs, over increasing time intervals of $\Delta t = \delta t, 5\delta t, 50\delta t,$ and $100\delta t$.

Chapter 5

Designing biological networks with ease: Network Designer

5.1 Overview and current difficulties in mathematical modelling

Mathematical modelling is a powerful tool that can be used to understand and explore the dynamics of biological systems. These models can be used to; make predictions about the behaviour of biological processes, infer experimentally inaccessible parameters, and test hypotheses about the underlying mechanisms that drive these systems. A wide variety of biological systems have been modelled mathematically, including population dynamics, neural networks, and gene regulatory networks.

In my studies I have focused on modelling how DNA, mRNA, and proteins move and interact with one another to generate circadian rhythms. Whilst trying to construct models of the circadian clock, I encountered a number of difficulties and inefficiencies. Chief amongst these issues was that large mathematical models became difficult to prototype quickly and cumbersome to alter. Simple models involving feedback mechanisms typically generates systems of coupled ordinary differential equations (ODEs) with tens to hundreds of subtly different terms. Thus, when writing down these equations incorrect terms and typo errors frequently occurred. Misdefining even one of these terms leads to software crashes and hard to spot false results. Hence, I sought to address these issues through my software *Network Designer*, by enabling quicker exploration of models through graphical manipulation of reaction networks combined with automated generation of differential and stochastic equations. Whilst this software took significant time to develop, it vastly reduced the time it took me to define new models. Adding or deleting nodes became trivial rather than an arduous process of checking through coupled equations. Network Designer was used throughout my work and was used to develop the models of DNA binding regulation seen in chapter 2 and facilitated protein-protein complexing in chapter 3. Throughout it's conception, I have been working with other researchers to improve the usability of Network Designer, with an aim to release the software to the wider research community.

Other software solutions, such as VCell, CellNetAnalyzer, and SBML Editor, exist for assisting with the mathematical modeling of biological systems (Blinov et al., 2017; Klamt et al., 2007; Hucka et al., 2003). I found that these programs did not provide the level of efficiency I was seeking when defining reaction equations. Additionally, the Kappa language also has a graphical user interface for defining models. However, Kappa focuses on rule-based models, rather than traditional mathematical models, thus is unsuitable for my work. As a result, I developed Network Designer to streamline my modelling workflow. Network Designer focuses on having a user-friendly interface to interact with a network graph that auto-generates the system of ODEs and rate equations. These features enables quick prototyping of new models and exploration of multiple models without requiring extensive programming or mathematical knowledge.

5.2 My solution: Network Designer

Network Designer, is a user-friendly computer application that enables researchers in the life-sciences to easily create, simulate, and fit mathematical models of biological systems. My goal was to provide a tool that allows users to quickly prototype new models in a matter of minutes, without the need for specialist programming or mathematical knowledge. By leveraging the power of open-source libraries such as SciPy (Virtanen et al., 2020), Numpy (Harris et al., 2020), StochPy for simulations (Maarleveld et al., 2013), Cytoscape for network visualisation (Shannon et al., 2003), and the Plotly Dash framework for a browser-based interface (Inc, 2015), Network Designer aims to provide a streamlined and accessible solution for creating, simulating, and analyzing mathematical models of biological systems (Figure 5.1). Additionally it is cross-platform compatible, and can be easily hosted on a web server. The main objectives of Network Designer are to;

1. Be user-friendly,
2. Allow models to closely reflect diagrams of biological system,
3. Reduce the need for writing and checking equations by hand,
4. Enable rapid exploration of multiple models.

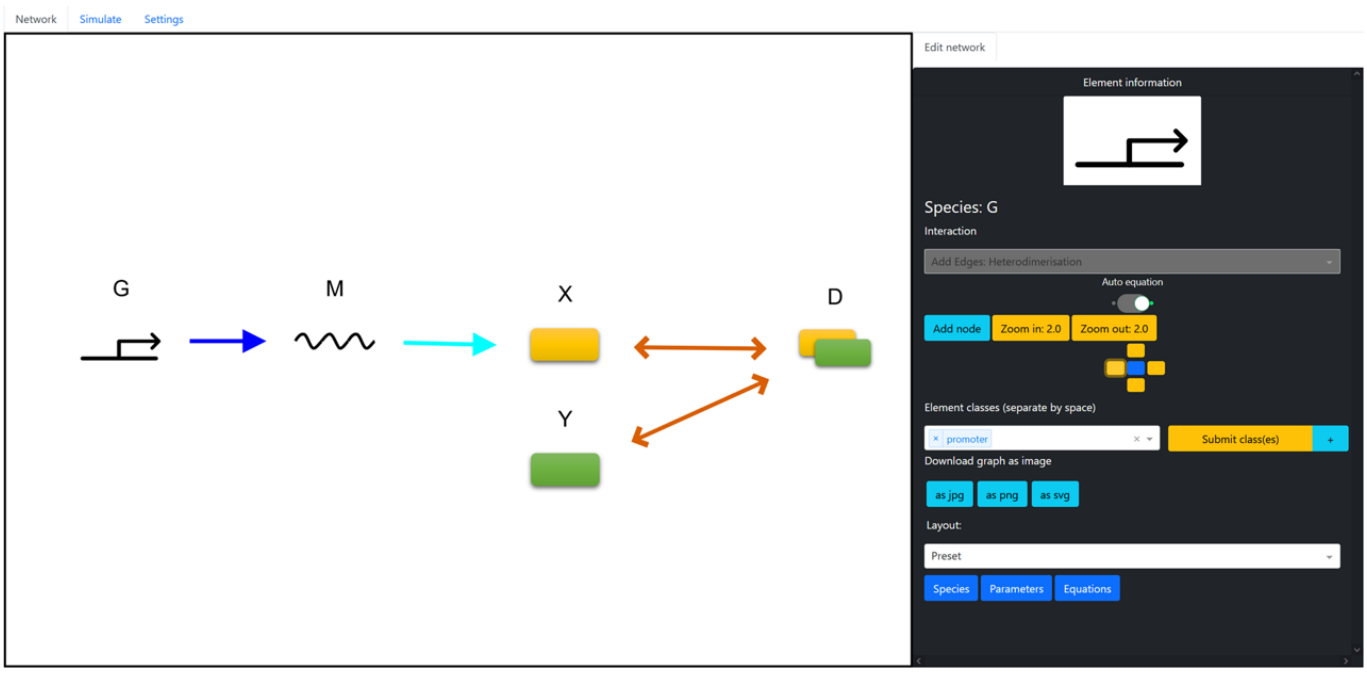

Figure 5.1: Screenshot of Network Designer. In Network Designer the user interacts with a Cytoscape network diagram to create mathematical models of biological processes. In the network tab of Network Designer the user can manipulate the graph of species and reactions using the dropdown menus and pop-up modals for altering species, parameters, and equations.

5.2.1 Software architecture

I implemented Network Designer in Plotly's Dash considering it was already built in Python and thus compatible with other modelling code I had already written. Furthermore, Dash is particularly suited to build a data dashboard application with a customised user interface. Network Designer operates with a user interface comprised of dropdown menus, buttons, and a intractable Cytoscape network graph all within a browser window (Figure 5.1). These interactable elements take user input and execute a series of call back functions. These call backs save/load data, update the network graph, generate equations, perform simulations, and produce plots. Figure 5.2 summarises the software architecture of Network Designer .

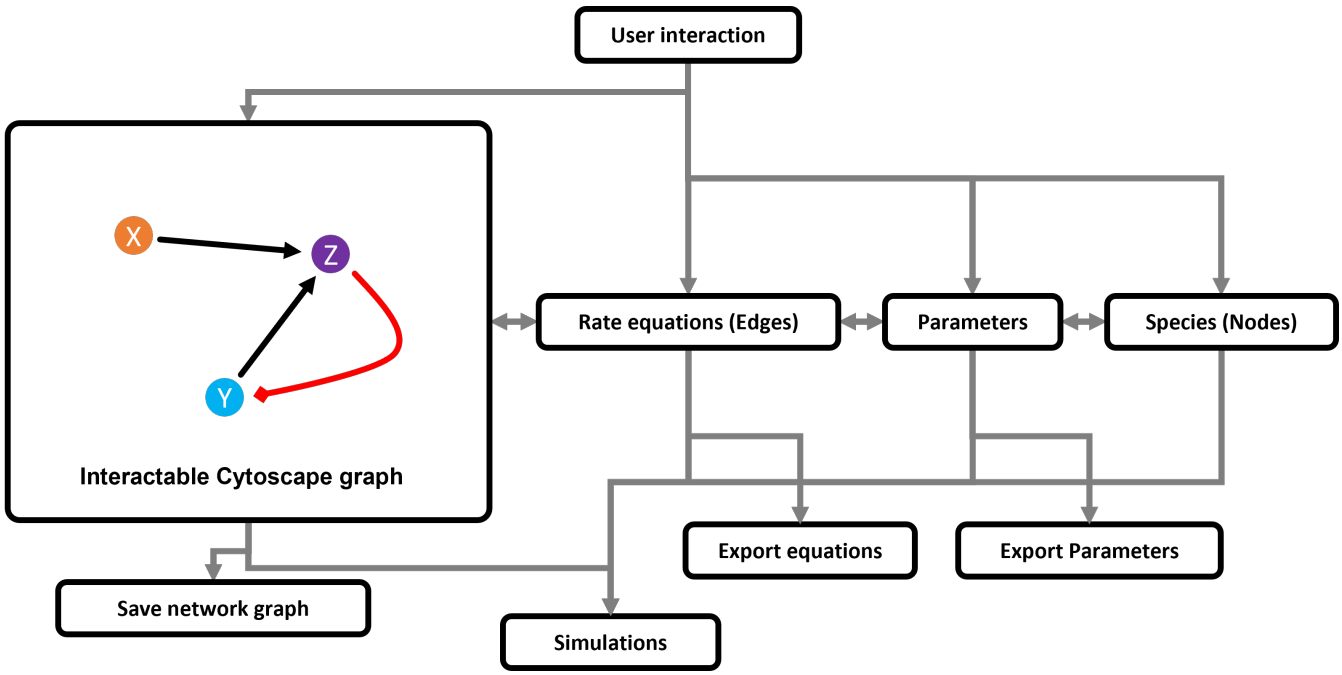

Figure 5.2: Diagram of the Network Designer Software Architecture. A visual representation of the software architecture for Network Designer, demonstrating how the user can interact with the Cytoscape graph and tables to define reactions, species, and parameters. These four elements are all interdependent, updating each other in real time as the user alters any parameters or features of the model. These inputs are then utilised to simulate the underlying model equations, save the network graph, as well as export equations and parameters. Arrows indicate overall flow of information between each element of the program.

The user interface (UI) of Network Designer was designed to be easy to use. I used CSS and HTML to customise the UI, making use of three tabs displayed at the top of the application to allow the user to toggle between the network, simulations, and settings. Within the network tab the user can manipulate all aspects of the model. I used pop-up modals for various functionalities, including saving and loading models, defining parameters, species names, and reactions. Interactable tables are provided so that reaction equations, parameters, and species can all be easily modified (Figure 5.3a). Within the parameters (Figure 5.3b) and equations modals (Figure 5.3c), users can import and export parameters as well as reactions respectively. Pictures for nodes are used to enhance the visual representation of the model in the Cytoscape network graph, with users having the option of uploading scalable vector graphics files (svg) to use as pictures for each node using the class system. In the settings tab, the user can customise how the ODE equations are displayed as well as toggle between light and dark display modes for the UI (Figure 5.4). Finally, in the simulations tab, the user can run different deterministic and stochastic simulations, with the total simulation time defined by the user within a table. Each simulation is given a unique identification number, making it possible to plot and compare different runs of the model. Plots of deterministic and stochastic simulations can be overlaid by the user utilising the interactable table and dropdown menu. In addition, histograms of variables over multiple stochastic trajectories are also displayed. The user also has the ability to upload data for the model to be fitted against, with the resulting parameters being stored and assigned a unique ID for later use. This allows the different fitted parameters to be applied to the model at a later time.

5.2.2 Key features of Network Designer

One of the key features of Network Designer is auto-equation generation, which enables the user to automatically generate mathematical equations based on the Cytoscape network diagram. As users manipulate the network by adding, removing, redefining, and renaming the nodes and edges, the corresponding equations and parameters also automatically update. Crucially, parameter names are

(a)

Species

Model name

default

Species

Species	Classses	Initial copy number	Concentration / M
G	promoter	2	1.11e-11
M	RNA	0	0
X	protein1	0	0
Y	protein2	1e+4	5.54e-8
D	protein_dimer	0	0

Volume / L

3e-13

(b)

Parameters

Parameters

Add parameter

Drag and Drop or Select Files

Export

	Parameter	Value	Units	Lower Bound	Upper Bound	Fitable
×	ar_XY	0.001	/nM /s	0.0001	0.005	Yes
×	KD_XY	25	nM	10	100	Yes

(c)

Equations

Reactions

Export

Reaction	Reaction description	Reaction type	Propensity
$pool > M		Transcription	k_tx*G
$pool > X		Translation	k_tl*M
X + Y > D		Heterodimerisation	ar_XY*X*Y
D > X + Y		Heterodisassociation	ar_XY*KD_XY*D

Ordinary differential equations

(1) $\frac{dM}{dt} = k_{tx} \cdot G$

(2) $\frac{dX}{dt} = k_{tl} \cdot M - ar_{XY} \cdot X \cdot Y + ar_{XY} \cdot KD_{XY} \cdot D$

(3) $\frac{dY}{dt} = -ar_{XY} \cdot X \cdot Y + ar_{XY} \cdot KD_{XY} \cdot D$

(4) $\frac{dD}{dt} = ar_{XY} \cdot X \cdot Y - ar_{XY} \cdot KD_{XY} \cdot D$

Figure 5.3: Pop-up modals are used to define species, parameters, and equations. Any changes of the model are automatically shared between the network graph and these modals containing tables. (**a**) The nodes of the network graph can be manipulated within the species modal via the table. Classes may be defined to change the pictures of each node as well as add a description of what it represents. (**b**) The pop-up for parameters is where the user can define the names of parameters, their values fitting conditions, as well as import/export parameters to and from excel files. (**c**) All edges in the network represent reactions and are represented within the table of the equations pop-up modal. The rates of reaction, also known as the propensity, are converted into ordinary differential equations formatted and displayed in LaTex.

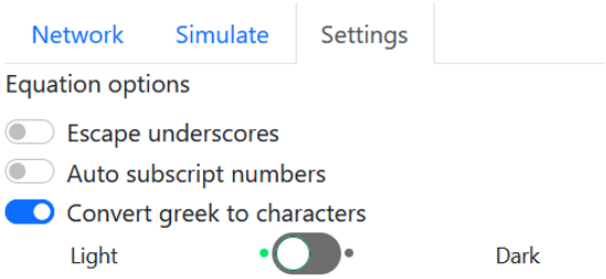

Figure 5.4: Settings tab. The user can define how ordinary differential equations are displayed in the settings tab as well as toggle between light and dark modes.

linked to the reactions and species they are part of to ensure no typos occur and that it is easier to attribute parameters to the variables they modify. Additionally, extra classes and descriptions can be added to each edge and node by the user so that the user can keep track of features of interest, making it easier to curate equations, molecular numbers, and parameters within one space. These automated features reduce errors and the need for specialist programming and mathematical knowledge, making it easier for researchers in the life-sciences to transform biological network diagrams into fully-fledged models. The equations used within the auto-equation feature are discussed in detail below.

Classes are also used to define pictures for each node, enhancing the ease of understanding. This makes the Cytoscape graph look like network diagrams often used by researchers to depict biological systems, see Figure 5.1. This network graph can then be exported as a picture in jpg, png, svg file formats. The equations can also be output to LaTex or Python to continue modelling outside of Network Designer.

Another important feature of Network Designer is simulating models both deterministically and stochastically. The user can easily convert between concentrations and copy numbers by defining the volume of the system, making it possible to simulate both deterministic and stochastic models simultaneously. The outputs of these simulations are then plotted in a variety of ways including line plots, average plots of stochastic simulations, and histograms of stochastic simulations.

Finally, users also have the ability to fit models to data using a genetic algorithm, with users able to fit both parameters and concentrations. To avoid fitting all initial concentrations and parameters, the user can define a subset of these that can also be fixed to certain values during fitting.

In Network Designer, each node of the graph represents the concentration of a molecular species and each edge a reaction between nodes. In addition to transformative reactions, each edge can also represent regulation of another reaction. For example, transcription factors and repressors can act as a catalyst or inhibitor to increase and decrease the rate of transcription respectively despite not be being used up in the overall reaction. Changes in the abundance of molecular species (nodes of the graph) can be modelled using an ordinary differential equation as a sum of its positive and negative reactions, represented as edges originating and terminating at the node. For a node representing species X the ODE of its rate of change is the sum over positive and negative reactions,

$$\frac{dX}{dt} = \sum \text{Positive reactions} - \sum \text{Negative reactions}. \tag{5.1}$$

Here, positive reactions are processes that increase the amount of X whereas negative reactions decrease X. It should be noted that nodes can represent concepts other than whole numbers of biomolecules and can instead represent concepts such as modifications to existing molecules like phosphorlyations or the number of co-factors bound to a promoter element of DNA. Let's consider a simple example of how Network Designer models the network graph for constant transcription and degradation of an mRNA, M, originating from the gene G which is then translated in a protein P

that is also degraded. The mRNA M evolves according to the ODE

$$\frac{dM}{dt} = k_{tx}G - k_{dm}M, \tag{5.2}$$

where $k_{tx}G$ is the positive rate of reaction and k_{dm} is the rate of degradation. The abundance of M increases according to the transcription rate k_{tx} and the concentration of the gene G, and decreases proportionally to its own abundance by the degradation rate $k_{dm}M$. The protein P proceeds this and its concentration is modelled as

$$\frac{dP}{dt} = k_{tl}M - k_{dp}P, \tag{5.3}$$

where k_{tl} and k_{dp} are the translation and degradation rates respectively. As you can see this ODE depends explicitly on M and thereby is coupled with equation (Equation 5.2) forming the complete system of equations. This is modelled in Network designer as a graph with nodes for the gene G, mRNA M, protein P, and the pool, $\emptyset$, with edges between each node (Figure 5.5).

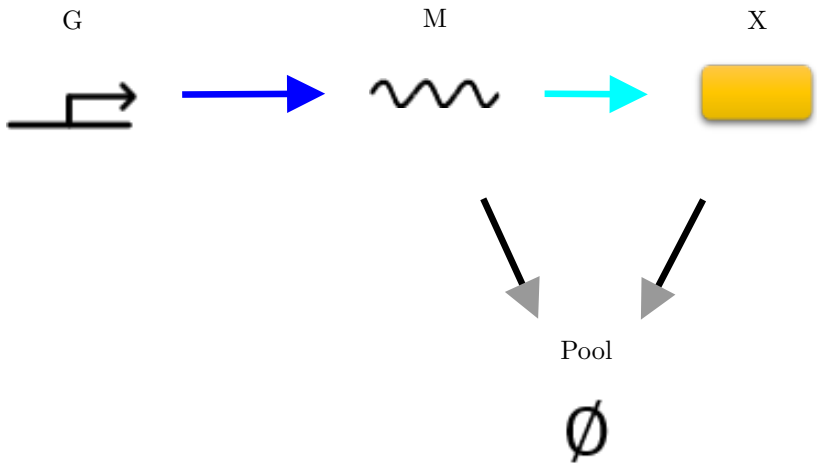

Figure 5.5: Simple transcription-translation model. A representative graph of a simple model of transcription of an mRNA, M, from a gene, G, that is subsequently translated into a protein P. Both the mRNA and protein are eventually degraded into a non-specific pool.

Here, the symbol, $\emptyset$, represents a pool from which species can be created from or degraded into. The pool is infinite and as such its concentration and copy number are not tracked. These edges representing with reactions can be simulated stochastically using Gillepies algorithm via the Python library *StochPy* by converting the ODEs to rate equations. These reactions are modelled individually and occur according to probabilities commonly referred to as propensities. The nodes are altered to explicitly track the copy number, here M, at each time-point and equations (5.2-5.3) are converted to the rate equations

Reaction	Propensity
$\emptyset \longrightarrow M$	$k_{tx}G$
$M \longrightarrow \emptyset$	$k_{dm}M$
$\emptyset \longrightarrow P$	$k_{tl}M$
$P \longrightarrow \emptyset$	$k_{dp}P$

M is created from the pool as the gene G here promotes the creation of M and is not used up in the reaction. The protein P is similarly created from and degraded into the pool.

Biological processes including transcription, translation, degradation, and dimerisation, can all be modelled using ODEs and rate equations. Often simple mass action kinetics are used, but more

complex Micheaelis-Menten type dynamics can also be used to represent activation (increasing rate) or repression (decreasing rate) dynamics dependent on the concentration of regulating factors. In Network Designer the reactions that each edge represent are defined by the user with the equations automatically assigned and given a class. The reactions currently implemented in the auto-equation feature of Network Designer are transformation, transcription, translation, degradation, and dimerisation. The rate parameters of each reaction are similarly automatically generated with different types of reactions/edges differentiated by colour. These reactions are discussed in detail below. In its current iteration Network Designer provides models of mass action kinetics using its auto-equation feature, with other models needing to be specified manually. However, working outside of the auto-equation feature is currently discouraged as consistency across the network cannot be guaranteed. In the future I hope to expand the number of biological processes that can be modelled using the automated features I have implemented within Network Designer.

Transformation

Transformation can encompass a variety of biochemical phenomena, including translocation between sub-cellular compartments, maturation of fluorophores, post-transcriptional modifications of RNA, and post-translational modifications of proteins (i.e. phosphorylation, ubiquitination, etc). Here, transformation is modelled as a process in which a molecular species is converted into a different species at constant forward and backwards rates, k_f and k_b respectively (Figure 5.6).

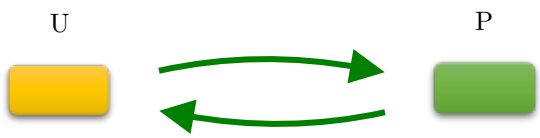

Figure 5.6: Transformation. Representative graph of a transformation process, in which molecule U is converted into P and vice versa.

When a transformation process occurs at a constant rate, the temporal evolution of a transformed molecule, denoted as P, can be accurately modelled using the coupled ODEs

$$\begin{aligned} \frac{dP}{dt} &= k_f U - k_b P, \\ \frac{dU}{dt} &= k_b P - k_f U, \end{aligned} \tag{5.4}$$

where U is the unprocessed molecule. Note that at equilibrium the forward rate of reaction is equal to the backwards rate. In certain biological systems, processes such as mRNA splicing and peptide cleavage, may not be able to be reversed and thus, the reverse reaction rate is considered to be zero. $k_b = 0$. The corresponding rate equations of a transformation process are auto-generated in Network Designer and can be seen in Figure 5.7.

Transcription

Transcription is the process by which the genetic information stored in DNA is converted into RNA (Figure 5.8). In mathematical modeling of transcription, various kinetic equations and rate laws are used to describe the rate at which the transcription process occurs, taking into consideration factors such as the affinity of transcription factors for the DNA, the availability of the RNA polymerase enzyme, and other molecular interactions. Within the auto-equation feature of Network Designer, transcription is simply modelled using mass action kinetics, where the rate of mRNA production, M, from a gene, G, is proportional to the concentrations of the gene and a rate constant, k_{tx}, as

$$\frac{dM}{dt} = k_{tx} G. \tag{5.5}$$

Equations

Reactions

Export

Reaction	Reaction description	Reaction type	Propensity
U > P	Forward reaction	Transformation	k_f*U
P > U	Backwards reaction	Transformation	k_b*P

Ordinary differential equations

(1) $\frac{dU}{dt} = -k_f \cdot U + k_b \cdot P$

(2) $\frac{dP}{dt} = k_f \cdot U - k_b \cdot P$

Figure 5.7: Transformation equations. Screenshot of automatically generated equations corresponding to (Equation 5.4) of a simple transformation process.

This equation is automatically converted into the rate equations by Network Designer (Figure 5.9).

Figure 5.8: Transcription Representative graph of transcription of an mRNA M from a gene G.

Equations

Reactions

Export

Reaction	Reaction description	Reaction type	Propensity
$pool > M		Transcription	k_tx*G

Ordinary differential equations

(1) $\frac{dM}{dt} = k_{tx} \cdot G$

Figure 5.9: Transcription equations. Screenshot of transcription equation auto-generated by Network Designer corresponding to Equation 5.5.

Translation

Translation is the process by which ribosomes translate messenger RNA (mRNA) into functional protein (Figure 5.10). Several biochemical reactions occur during translation, including initiation, elongation, and termination.

Figure 5.10: Translation. A graph of mRNA, M, translated into a protein P.

To describe the underlying kinetic process, simple mass action kinetics is often used, where the rate

of reaction is proportional to the concentration of the reactants. This is the approach used by the auto-equation feature in Network Designer. In this, translation from mRNA, M, into protein P can be modeled without considering the consumption of the mRNA in the reaction because one mRNA molecule can be translated multiple times such that the amount of protein produced in the cell is typically much higher than the amount of mRNA, meaning that the mRNA concentration can be considered to remain constant during the translation process. This approach allows for a simplified mathematical representation of the translation process, where the change in protein concentration is determined by the rate of translation, rather than the rate of mRNA consumption. Thus the constant rate of production of a protein P via translation of mRNA, M, at a rate k_{tl} can be modelled as the ODE

$$\frac{dP}{dt} = k_{tl}M. \tag{5.6}$$

My application converts these equations into rate equations as can be seen in Figure 5.11.

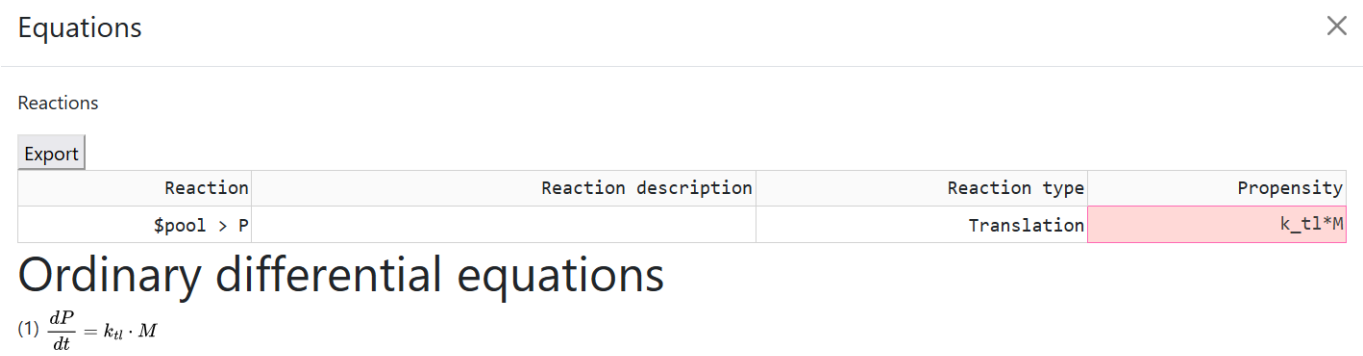

Figure 5.11: Translation equations. Screenshot of Network designer generated equation for an mRNA M being translated into protein P, corresponding to Equation 5.6.

Degradation

Degradation is the process of breakdown or removal of proteins, RNA, and other biomolecules (Figure 5.12).

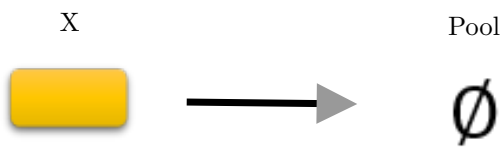

Figure 5.12: Degradation. Graph of the degradation of molecule X into a non-specific pool $\emptyset$.

This process is crucial in the regulation of cellular processes, such as regulating the abundance of cellular components, including proteins, lipids, and nucleic acids. The degradation process helps maintain the balance of these components, allowing cells to respond to internal and external signals and adapt to changing conditions. Additionally, degradation helps ensure that damaged or misfolded proteins are removed from the cell, preventing them from causing further harm. Degradation is especially important in an oscillatory system like the circadian clock to relieve repression to reinitiate the active phase of transcription. In cells, degradation depends on factors such as enzymatic degradation, autophagy, and other processes that lead to the removal of biomolecules. In Network Designer degradation is modelled as a first order process whereby the constant degradation of a molecule X

at rate k_d can be modelled as the ODE

$$\frac{dX}{dt} = -k_d X. \tag{5.7}$$

This leads to exponential decay of the molecule X, with larger numbers of X increasing the rate of degradation. Network Designer automatically converts this equation into the rate equations in Figure 5.13.

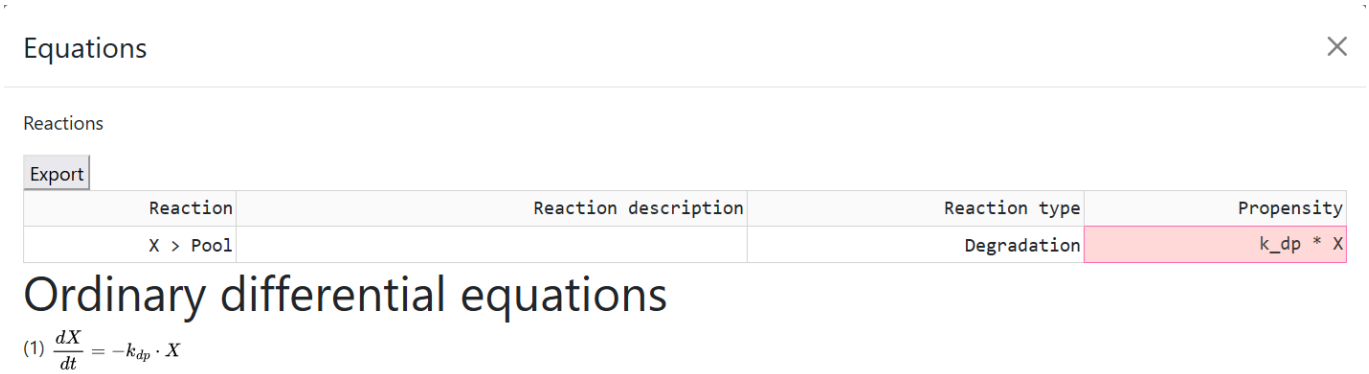

Figure 5.13: Degradation equations. Screen shot of auto-generated equations for the degradation of a protein X at rate k_{dp}.

Dimerisation

Dimerization is the process by which two separate molecules bind together to form a dimer, a single unit consisting of two identical or non-identical subunits. This process is a crucial step in many biological processes and is involved in a variety of physiological functions, including protein-protein interactions, signal transduction, and gene regulation. Dimerization can occur between identical molecules, known as homodimerization, or between different molecules, known as heterodimerization (Figure 5.14a). Heterodimerisation can be modelled using ODEs as

$$\begin{aligned} \frac{dD}{dt} &= k_a XY - k_d D, \\ \frac{dX}{dt} &= k_d D - k_a XY, \\ \frac{dY}{dt} &= k_d D - k_a XY, \end{aligned} \tag{5.8}$$

where k_a is the rate of association and k_d is the rate of dissociation. These rates are often hard to individually measure within experiments. Thus, the dissociation constant, K_D, is commonly used to represent the strength of dimerisation and is given by

$$K_D = \frac{k_d}{k_a} = \frac{XY}{D}. \tag{5.9}$$

K_D is an inverse measure of affinity and represents the concentration of X and Y at which 50% of each will be bound within the dimer D. The dissociation constant is typically the value reported for protein-protein interactions within literature and is quantified by techniques such as fluorescence cross-correlation spectroscopy (Bacia et al., 2006). In Network Designer, k_d is replaced by the K_D

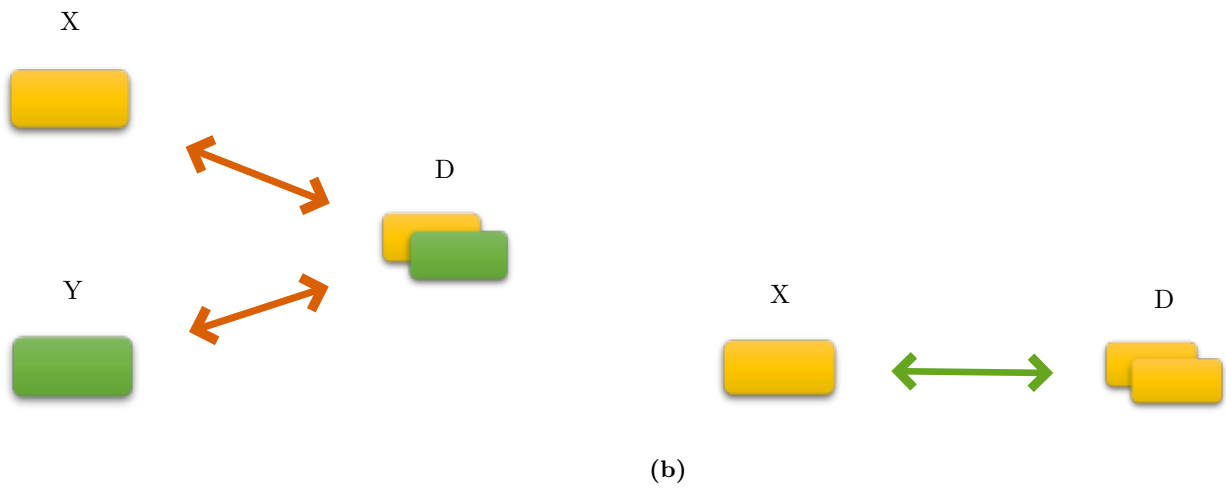

Figure 5.14: Dimerisation. Graphs of (**a**) X and Y heterodimerising to from a dimer D and (**b**) of two X molecules homodimerising into the homodimer D.

and the association rate using Equation 5.9 to convert Equation 5.8 into

$$\begin{aligned} \frac{dD}{dt} &= k_a(XY - K_D D), \\ \frac{dX}{dt} &= k_a(K_D D - XY), \\ \frac{dY}{dt} &= k_a(K_D D - XY). \end{aligned} \tag{5.10}$$

This framework is especially useful when modelling the steady state equilibrium of dimerisation as knowing the true value of k_a is not required. Instead k_a now models how quickly the dimer forms and not the fraction of binding. Thus, the ratio of monomer to dimer is purely modelled by the K_D. The automatically generated rate equations for dimerisation are given in Figure 5.15.

Equations

Reactions

Export

Reaction	Reaction description	Reaction type	Propensity
X + Y > D		Heterodimerisation	ar_{XY}*X*Y
D > X + Y		Heterodisassociation	ar_{XY}*K_{D}(X,Y)*D

Ordinary differential equations

(1) $\frac{dX}{dt} = -ar_{XY} \cdot X \cdot Y + ar_{XY} \cdot K_D(X,Y) \cdot D$

(2) $\frac{dY}{dt} = -ar_{XY} \cdot X \cdot Y + ar_{XY} \cdot K_D(X,Y) \cdot D$

(3) $\frac{dD}{dt} = ar_{XY} \cdot X \cdot Y - ar_{XY} \cdot K_D(X,Y) \cdot D$

Figure 5.15: Heterodimerisation equations. Screenshot of auto-generated heterodimerisation equations corresponding to Equation 5.10.

In contrast to heterodimers (Figure 5.14b), if a molecular species can bind itself it may form homod-

imers according to the rate

$$\begin{aligned} \frac{dD_{\text{Homo}}}{dt} &= k_a X - k_d D_{\text{Homo}}, \\ \frac{dX}{dt} &= k_d D_{\text{Homo}} - k_a X. \end{aligned} \tag{5.11}$$

These ODEs can be transformed into the rate equations in Equation 5.2.2. Larger networks of multimeric protein complex formation can be modelled using a series of these dimerisation reactions.

Reaction	Propensity
$2X \longrightarrow D_{\text{Homo}}$	$k_a X$
$D_{\text{Homo}} \longrightarrow 2X$	$k_d D_{\text{Homo}}$

Table 5.1: Rate equations for homodimerisation of two molecules X into the homodimer D_{Homo}.

5.2.3 Numerical simulation and fitting

In network designer, user-defined models are simultaneously represented as differential equations and rate equations, hence both deterministic and stochastic numerical simulations can be performed. Once the model has been defined, network designdefineder can numerically simulate the model over a specific time interval from an initial set of concentrations. The user can perform multiple simulations over different combinations of parameters, allowing them to compare the results of these simulations. This enables the user to explore the impact of different parameter combinations on the behaviour of the system. For example, different combinations of parameter values can be used to represent different biological conditions, or to test different hypotheses about the system. After each simulation has been performed, network designer saves the results, allowing the user to easily compare the results of multiple simulations (Figure 5.16).

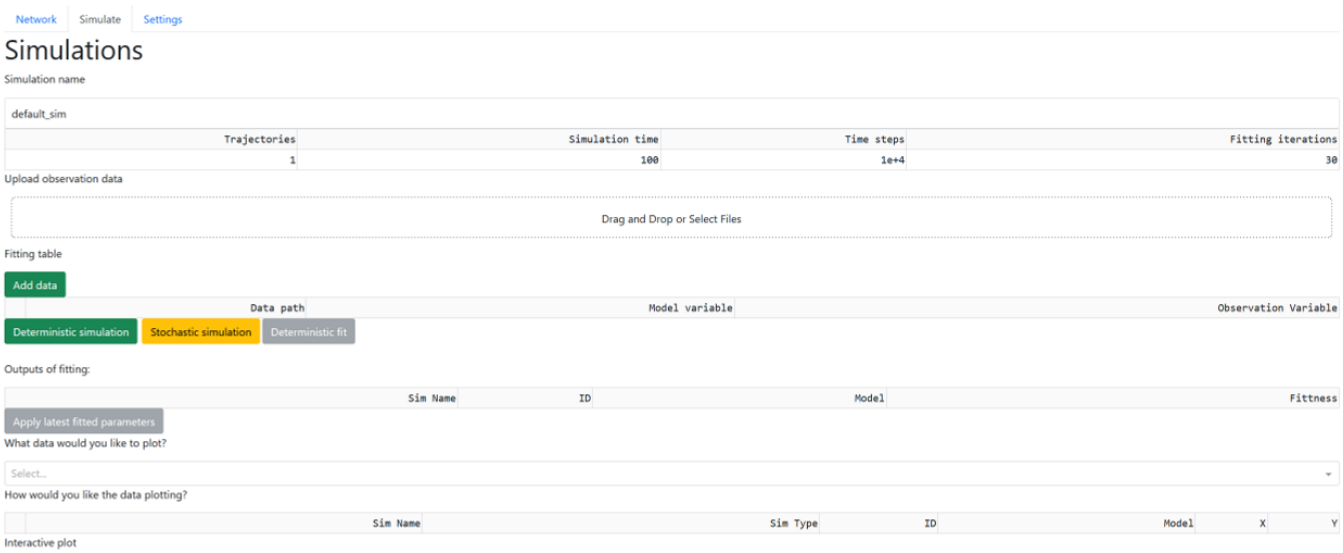

Figure 5.16: Simulation tab. Screenshot of simulations tab in Network Designer where the user can run deterministic and stochastic simulations as well as fit models to data. Users can name simulations for later referencing and plotting. Users upload data to fit the model via the 'Add data' button which is then added to the table below, the user can then select which variables of the model that the data represents.

Simulated over a specified time interval, the deterministic simulation process involves solving the ordinary differential equations (ODEs) that represent the model and calculating the state of the system at each time point within the defined interval. This is accomplished through the use of the odeint function from the Python package SciPy.

Stochastic simulations are achieved by transforming the deterministic ODE model into a set of rate equations, which can then be used to make a '.psc' file for stochastic simulations performed by the Python package StochPy. In these simulations, random fluctuations are introduced into the system, reflecting the inherent variability and uncertainty in biological processes. The stochastic simulation algorithm generates multiple simulation runs, or trajectories, which can then be averaged to produce an ensemble of simulations. This approach provides a more comprehensive and realistic representation of biological systems compared to deterministic simulations. This is because it accounts for the inherent variability and stochasticity in biological processes. Furthermore, the ability to generate multiple simulation runs and average the results allows the user to account for the effects of random fluctuations on the overall behaviour of the system. Overall, the integration of StochPy with Network Designer offers a powerful tool for simulating and analyzing biological systems, providing a more complete and realistic representation of biological processes and their underlying mechanisms.

Fitting to data

In Network Designer, deterministic ordinary differential equation (ODE) models are fit using the genetic algorithm *differential_evolution* from the SciPy Python package. The goal of this process is to find the best set of model parameters that fit the data and capture the underlying biological processes being modeled. The *differential_evolution* algorithm uses a population-based search, where a group of candidate solutions is iteratively evolved to minimize the discrepancy between the model predictions and the data. Users have the option to specify certain parameters of the model to be fixed at a user-defined value, which can help constrain the solution space and improve the fitting process. This is particularly useful when certain parameters are known *a priori* or when a model has multiple local minima, and fixing certain parameters can help prevent the optimisation algorithm from being trapped in a sub-optimal solution. Users can input multiple data sets corresponding to different parameters of the model, such as smFISH data for mRNA counts and FCS concentration data for protein abundance, all within a transcription-translation model.

5.3 Discussion and future work

I developed Network Designer with the goal of streamlining the mathematical modeling process in biological systems by curating models, data, and experiments into one application. I would like to further develop Network Designer for gathering data, exploring model topologies, testing hypotheses with data, and outputting useful simulations, values, and diagrams. It is my hope that Network Designer will in the future prove to be a valuable tool for researchers in the field of biological modelling. To support these aims, I would like to expand Network Designer by incorporating a number of new features to the software to improve its functionality and usability. These features are:

1. **Expand auto-generated equations.** I would like to expand the equations offered by the auto-equation feature beyond mass-action kinetics to include more complex dynamics like Michaelis-Menten terms.

2. **In silico experiments.** Increasing the variety of simulations that can be performed would allow users to perform in silico experiments as well as further analyse their models. For example, sensitivity analysis could highlight the most sensitive points of the model. Furthermore, models could be better linked to experiments by adding special nodes that represent processes such as observing photon counts emitted from a fluorophore or averaging over many cells during end-point sample collection.

3. **Additional fitting options.** Currently, Network Designer only allows for fitting of parameters and concentrations by a genetic algorithm. I want to include other types of modelling and fitting

that better account for uncertainty, such as likelihood models, chemical Langevin equations, and Monte-Carlo simulations.

4. **Library of existing models.** I wish to integrate Network Designer into a library of commonly used and previously defined models such as the BioModels database (Malik-Sheriff et al., 2020). This would allow users to quickly access previously defined models and reduce the time spent on model construction.

5. **Define models using natural language.** I wish to incorporate natural language model definition using something like ChatGPT, a natural language model by OpenAI. This would allow users to define models in words, using more intuitive and natural language. Thus making it easier to understand and construct models.

6. **Version history.** Version history using a system like Git, so that users can see branch points and keep track of changes made to their models.

7. **Check identifiablity.** Include automatic identifiablity checking of models. This would examine whether certain components of the model are redundant and if they can be appropriately defined during fitting. This could help users identify if it is possible to infer certain parameters and if the model outstrips their data.

8. **Integrate with existing frameworks.** Integrate Network Designer within existing frameworks, such as by importing/exporting to and from the Systems Biology Markup Language (SBML) (Hucka et al., 2003; Finney and Hucka, 2003). This would allow users to apply their models in other applications, and allow for better collaboration and sharing of models.

9. **Filter and manipulate by classes.** Allowing the user to filter and view nodes/edges of the network by classes would to improve organisation and management of models.

10. **Sub-cellular compartments.** Add the ability to explicitly define different sub-cellular localisations via compartments such as the nucleus and cytoplasm. This would allow the users to model how molecules interact within specific sub-cellular regions and how they traffic between compartments.

11. **Lookup bio-molecular data.** It would be highly useful if users should search large data repositories for data on the biomolecules they are building models from. Amongst others, these databases could include UniProt for protein data (The UniProt Consortium, 2022), AlphaFold for protein structure (Jumper et al., 2021; Varadi et al., 2021), and National Center for Biotechnology Information (NCBI) for other databases (Sayers et al., 2021).

Chapter 6

Discussion

6.1 Summary

In this thesis, I have established a new understanding of the operation of the circadian clock. I achieved this by developing novel quantitative approaches, which enabled modelling of the molecular mechanisms responsible for generating circadian rhythms in cells. Quantitative microscopy techniques, including Fluorescence Recovery After Photobleaching (FRAP) and Fluorescence Correlation Spectroscopy (FCS), were used to study these fundamental mechanisms. I constructed mathematical models to further explore the data, shed light on experimentally inaccessible complexes, and elucidate the repressing mechanisms of the key transcription factor CLOCK:BMAL1. I tested these experimentally. To support these extensive FCS measurements, I created a bespoke analysis pipeline to improve inferences and reduce analysis time from hours to minutes. A new methodology, FCS-Maximum Approximate Likelihood (FCS-MAL), was introduced to address the issues prevalent in traditional FCS analysis and to infer the same values with up to 1000 times less data. Finally, I created a new graphical modelling tool, Network Designer, to enable the construction of mathematical models of biological systems in an intuitive way, which reduced the human error in model construction.

6.2 Quantifying and modelling the circadian clock

6.2.1 Mechanisms of CLOCK:BMAL1 DNA binding

The molecular circadian clock drives the expression of a large number of genes through binding of CLOCK:BMAL1 to E-Box sites. This imposes a time-signature on an extraordinary number of biological processes (Koike et al., 2012). Using FRAP (chapter 2), I demonstrated daily rhythmic binding of the CLOCK:BMAL1 transcription factor to a range of DNA binding sites, which accorded well with previous findings from chromatin immunoprecipitation sequencing studies (ChIP-seq) (Koike et al., 2012). I showed that CLOCK is a cytoplasmic protein, and that BMAL1 regulates CLOCK nuclear entry at a 2:1 ratio through direct dimerisation. Individually these proteins are unable to effectively bind DNA. However once incorporated as a CLOCK:BMAL1 dimer they resided on DNA for an average of $\sim$ 3 s. This is in accordance with other known transcription factors such a GR, p53, p65, and STAT1 (Hettich and Gebhardt, 2018). This residence represents an average time of CLOCK:BMAL1 on DNA as FRAP is a bulk measurement of 10-50 % of the nucleus. Crucially, it is shorter than the $\sim$ 8 s residence time measured for a set of short strands of the DBP gene (Stratmann et al., 2012). Thus, the measurements made here accord well with the distribution of peak amplitudes in the ChIP-seq data that places almost all genes bound by CLOCK:BMAL1 less strongly bound than DBP (Koike et al., 2012). This short 3 s binding time, is hypothesised to reduce noise in gene expression. In combination with FCS/FCCS measurements of concentrations, diffusion, and inter-

actions, I then constructed a mathematical model to link the PER2:CRY1 complex as the regulator of this rhythmic DNA binding and showed that the PER2:CRY1 complex binds to CLOCK:BMAL1 to alter its affinity for DNA. This mechanism reduces the global occupancy of CLOCK:BMAL1 on target E-box DNA sites. However, modelling also revealed that, despite the reduction in overall binding, this mechanism also increases the frequency of CLOCK:BMAL1 visits to target sites. Thus, PER2 acts both as a repressor and indirectly as an activator. Intriguingly, CLOCK:BMAL1 was found to cycle between 500 and 1500 available dimers over the course of a circadian day, far fewer than the estimated number of tissue-dependent 2,000 to 6,000 high-confidence CLOCK:BMAL1 DNA targets identified by ChIP-seq (Rey et al., 2011; Koike et al., 2012; Wu et al., 2017; Oishi et al., 2017; Beytebiere et al., 2019; Chiou et al., 2016; Dyar et al., 2018). As a result, in order to perform its role as a transcription factor on these sites, CLOCK:BMAL1 must be a highly mobile transcription factor. Modelling further revealed that the changing DNA binding capabilities of CLOCK:BMAL1 through its binding with PER2:CRY1 performs a dual role: increasing the rate of DNA unbinding reduces the global number of bound DNA sites, while increasing the rate at which CLOCK:BMAL1 visits these sites. As such, CLOCK:BMAL1, along with site-specific binding profiles and co-factors, is capable of performing rhythmic transcriptional control that is out of phase with its global binding kinetics. This is consistent with a "hit-and-run" model of frequent short duration DNA binding by transcription factors (Shah et al., 2019).

I also provided the first direct measures of CRY1 protein binding affinity in living cells, finding that this protein forms a strong complex with BMAL1 and hence presumably the dimer CLOCK:BMAL1. This fits well with substantive evidence for direct repression of BMAL1 transactivation capabilities by CRY1 (Gustafson et al., 2017; Xu et al., 2015). In this study I showed that this interaction is not only strong but also rhythmic. Strikingly, CRY1 exhibits near-persistent binding with BMAL1 over 24 h owing to rhythmic interaction strengths countering changes in abundance. However, the exact mechanism that underlies this modulation of protein-protein interaction is yet to be determined. Binding with third partners, such as PER2, or post-translation modifications could explain the change in BMAL1-CRY1 affinity (Ye et al., 2011; Fribourgh et al., 2020; Schmalen et al., 2014).

In summary, this study examined the rhythmic interactions of CLOCK:BMAL1 for DNA as well as cycling protein-protein interactions of CRY1 for BMAL1. Mathematical modelling further explored the experimentally inaccessible complexes and revealed mechanisms for regulating this rhythmic DNA binding.

6.2.2 Protein-protein interactions in the circadian clock

I then used FCS/FCCS to examine the interaction between PER2 and CRY1 in a double knock-in mouse, finding that the interaction between PER2 and CRY1 was strong, rhythmic, and closely followed the 24-hour dynamics of the BMAL1-CRY1 interaction (chapter 3). To further expand the understanding of the interactions between key circadian clock proteins, I went on to study 15 different protein complexes comprising CLOCK, BMAL1, PER2, CRY1/2, and CK1δ, using a lentiviral expression system in fibroblast cell lines. Such an abitious clock-wide comparison has never been undertaken before. My findings showed that the majority of these protein-protein interactions were conserved across different cell types and protein expression levels.

By using mathematical modelling, I compared the suitability of a direct binding model or a three-way facilitatory binding arrangement to explain the data, and found that most proteins in the circadian network directly bind to each other in a pair-wise fashion with only a few key interactions facilitated by a third partner. These latter cases included CRY1/2 binding with CLOCK via both BMAL1 and PER2, as well as CK1δ binding CLOCK via PER2. This modelling placed BMAL1 and PER2 as central players in the network of protein-protein interactions in the circadian clock, likely acting as scaffolds for further multimeric complex formation.

These results were further experimentally validated through FCCS measurements of three-way interactions, which provided evidence of how circadian proteins reciprocally regulate each other's

protein abundance to maintain a stoichiometric balance. An outcome of this study was the demonstration that high affinity interactions between PER2 and CRY2 were sufficient both to stabilise CRY2 proteins and maintain post-translational rhythms of CRY2 independently of transcription. This offers novel insights into the dynamic interactions between various circadian protein components within living cells and underscores the crucial role that key interacting partners play in the circadian clock network.

This study could be extended in several directions, the most straight forward would be to expand the set of pairwise interaction measurements to other proteins in the circadian clock network such as the interaction of FBXL3 with CRY1/2, CK1ϵ with PER1/2, and BMAL1 with NPAS2. Additionally, exploring interactions of the core circadian components with proteins from other molecular systems in this manner is crucial to understanding cross-talk between these systems and the clock. For example, it is known that BMAL1 is able to dimerise with other Per-Arnt-Sim (PAS)-domain-containing proteins that form basic Helix-Loop-Helix transcription factors such as the oxygen sensing Hypoxia Inducible Factor (HIF) and metabolic regulator Aryl hydrocarbon Receptor (AhR) (Peek et al., 2017; Jaeger and Tischkau, 2016). The strength of these interactions are yet to be fully defined in living cells. The three-way protein expression data in this study clearly demonstrates the important of complex formation for additional post-translational modulation of circadian protein abundance. Prior studies have shown that PER2 can complex with and protect CRY2 from FBXL3 mediated degradation (Nangle et al., 2014). However, we also saw CRY1/2 'stabilise' following co-expression of BMAL1. This potentially novel mechanism of BMAL1 stabilisation of CRY1/2 requires further experimental validation. For example, measuring protein abundances after treatment with cycloheximide that are singly or co-expressed alongside partners should demonstrate whether or not these partnerships protect against degradation. Additionally, some of these two, and possibly three-way interactions, should be measured in the cells which lack other interacting proteins, i.e. measuring the interaction of the PERs for CLOCK and BMAL1 in CRY1 and CRY2 deficient cells. The issue of identifying and quantifying multivalent interactions between three or more proteins at once in live cells is non-trivial. In this study we focused on a small subset of possible multivalent interactions which could be expanded upon through combinations of pairwise FCCS in triple or even quadruple lentiviral over expression regimes both on wild type and different gene knock-out backgrounds. Additionally, FCCS may be adapted to identify possible multivalent interactions as the data should curve above or below the straight direct line if there is a required intermediate binding partner or competition respectively.

6.3 Analysis of fluorescence correlation spectroscopy data

6.3.1 Improving traditional FCS analysis pipelines

To support the extensive use of FCS and FCCS in studies of CLOCK:BMAL1-DNA binding (chapter 2) and protein-protein interactions (chapter 3), I developed a bespoke FCS analysis pipeline that addresses the problems of arbitrary analysis choices, excessive user interactions, and software issues prevalent in existing software such as PyCorrFit (Müller et al., 2014) and Fluctuation Analyzer 4G (Wachsmuth et al., 2015). The pipeline, described in the Materials and Methods section of chapter 2, reads and analyses multiple FCS files at once from a directory, streamlining the analysis process.

I created this custom Python software to enable me to experiment and refine the FCS analysis for my data sets. To avoid the possibility of fitting local minima and obtaining incorrect values, I implemented an initial round of fitting using a genetic algorithm, followed by a least squares minimisation algorithm to obtain uncertainty estimates for the fitted parameters. Additionally, I applied a model of systematic technical uncertainty on the auto- and cross-correlation curves arising from molecular processes, as described by Saffarian and Elson (2003). This model accounted for increased variability in the correlation at long and short lag-times, and made the least squares

regression less sensitive to high variance at the extremes of long and short lag-times, without resorting to the common practice of removing these high variance data.

Since inference techniques can only be as accurate as the data used, I removed outliers in the data sets that would arise from poor FCS measurements caused by out of focus light or low photon counts. These outliers can result in the correlation values being smaller than the uncertainty values, leading to unphysical values for concentrations and diffusion rates. With this bulk approach, I can now fit multiple models simultaneously to each correlation curve, and assign an Akaike Information Criterion Score (AIC) to each plausible model for each FCS experiment. This allows me to determine the optimal model, as well as explore changes in protein behaviour, such as how a protein changes from normal diffusion to anomalous diffusion when co-expressed with its binding partner.

This pipeline underwent changes between chapter 2 and chapter 3, as robust inferences of interaction strength from FCCS measurements were challenging. To determine the dissociation constant, K_D, in FCCS between partners A and B, good fits are required for both auto-correlation and cross-correlation channels. If any of these three fits are incorrect, the linear formula $K_D = [A]_{\text{Free}}[B]_{\text{Free}}/[AB]$ becomes corrupted. This is particularly problematic for cross-correlations, as we are comparing both signals. Unless the signals are strongly correlated due to strong complexing between the partners, the cross-correlation will be low and comparable to noise across most lag-times, often resulting in poor fits with a very low concentration for the complex $[AB]$. When the concentration $[AB]$ is small, the fraction comprising K_D becomes very large, reducing the apparent interaction strength.

In my first paper Koch et al. (2022) (chapter 2), these issues were mostly avoided due to the strong interactions between CLOCK-BMAL1 and PER2-CRY1 partners, as well as the hundreds of measurements taken. In this study, the K_D was also estimated by fitting the non-linear equation $[AB]/[A] = [B]_{\text{Free}}/(K_D + [B]_{\text{Free}})$. However, when investigating the many combinations of partners in chapter 3, this approach became impractical, as interaction affinity varied widely and screening poor measurements became very time-consuming. Additionally, non-linear fitting underestimated the uncertainty on fitted parameters too much and did not properly capture the true variance in K_D that was apparent when examining the plots.

I therefore switched to a linear fit, $[AB] = K_D[A]_{\text{free}}[B]_{\text{free}}$, and replaced the non-linear least squares regression with a robust linear regression through a quantile regressor (Koenker and Hallock, 2001). This type of robust regression is insensitive to outliers in the data, as it fits the inter-quartile ranges and median of the data, systematically ignoring parts of the data. Outlier data that exists outside of the middle 50% of data makes no contribution to the median estimate. I used this regressor to estimate the median slope of the line to determine K_D and its standard deviation. Quantile regression enabled the simultaneous robust estimation of K_D values and concentrations for multiple interacting partners. This allowed me to explore three-way interactions by changes in K_D between a pair of red and green proteins as a third blue partner was introduced. This change to a linear model and use of robust regression in chapter 3 led to lower measured K_D values, representing stronger complex formation, compared to those originally measured in chapter 2. Despite this discrepancy in absolute strength, the relative relationships of interaction strengths was preserved across both studies.

This pipeline currently exists as an executable Python script without a user interface. In the future, it could be adapted into a more user-friendly form by incorporating a graphical user interface. Additionally, the software could be easily adapted to work with a wider range of data from different microscopes, as it is currently limited to analysing Zeiss '.fcs' file types. It may also serve as a template from which alternative FCS analyses can be developed to analyse bulk FCS data.

6.3.2 New approaches

Throughout these two experimental studies, I observed that FCS was particularly prone to incorrect fitting of data, especially when examining low abundance fluorescent proteins. This issue was partially

alleviated by measuring more cells for longer and carefully adjusting laser intensities to prevent bleaching. However, poor fitting remained prevalent, requiring robust removal of outliers following fitting. The cause of these issues is that time correlating the data using multi-tau correlators leads to a large compression of data, from around $10^5 - 10^6$ points down to $\sim 10^2$ data points with widely varying uncertainties. Least squares fitting to these ~ 100 data therefore becomes very sensitive to a few outliers, making it unreliable even when measuring for longer. This highlights the need for an alternative FCS solution. Ideally, gathering more data points should improve inferences. Both principled Bayesian and maximum likelihood (a frequentist approach) statistical techniques meet this criterion.

I started with computational simulations of an FCS experiment to create synthetic data, by modelling N particles diffusing in 3D space according to standard normal distribution. The rate of photon emission was then calculated from the summation of the position of the N particles relative to the origin, the optical properties of the system and the per molecule emission rate. I examined how varying concentrations and diffusion rates influenced the distributions of intermediate parameters, such as the point spread function and emission rates of photons. I found that these distributions were well-behaved and could be predictably parameterised by both molecular abundance and diffusion rate. However, in this initial analysis, these parameters were related to summary statistics of the data, such as the mean and variance of the changes in photon counts over time. While this was an important step forward, these methods for inferring concentrations and diffusion rates still relied on binning the raw photon arrival time data into time bins, making them not reliable enough.

To address these issues, I sought to find a principled solution by deriving a relationship between the parameters we wished to infer and the set of raw, unprocessed photon arrival time data, $\underline{\Delta t}$. I began by calculating an approximate likelihood of observing the differences in arrival times, parameterised by the molecular brightness, concentration, and diffusion rate. By maximising this approximate likelihood, denoted as FCS-MAL, I succeeded in creating estimators that can accurately infer concentration and diffusion rates from FCS data, using only a few milliseconds of data rather than several seconds. These estimators relate the concentration rates to the average differences in photon arrival times, $\langle \underline{\Delta t} \rangle$, and the sum of covariances between these arrivals, $\sum_j \mathrm{Cov}(\Delta t_i, \Delta t_{i+j})$, respectively. This new methodology combines the computational efficiency of traditional FCS with the accuracy and reduced data requirements of recent Bayesian non-parametric techniques introduced by Jazani et al. (2019) and Tavakoli et al. (2020).

Further work is required to fully complete my model prior to publication. Specifically, the estimator for the diffusion rate has not yet been fully mathematically derived. As a result, the constants relating the covariance over Δt values to the diffusion rate must currently be found through calibration and fitting. This is down to the complexity of the model and the order of integrals I have chosen to complete first. Due to the nature of summing up all the contributions from each of the N molecules into one emission rate, an unavoidable mixing occurs. This leads to N^K possible combinations, for K possible photon arrival times, of subtly different integrals, whose results must be coherently summed up within the final likelihood. As such, the work presented in this thesis serves as a proof of concept for FCS-MAL until a complete expression for the diffusion estimator can be derived. Furthermore, the uncertainties associated with these estimators have yet to be defined, and methods for doing so are currently being sought. With collaborators, I intend to redo this work within a new mathematical framework that uses matrix representations and applies the Laplace approximation at a later stage in the calculation of the model likelihood. This representation allows us to greatly simplify the calculation as we can make use of established techniques for computing logarithms of matrix determinants. This will provide a more complete and accurate representation of the model.

Alternatively, this diffusive system could have been represented using a birth/death process as described by (Davis, 1965). In this approach, particles enter and exit the confocal volume based on both the diffusion rate and the current number of particles within the volume. However, a significant consideration is that the probability of detecting each particle is influenced by its spatial position

relative to the confocal volume's center. To accurately simulate this system, a spatial position grid would be required, introducing complexity to both simulation and subsequent inference. Instead, I chose to adopt the modeling approach delineated by Jazani et al. (2019) and Tavakoli et al. (2020). By directly modeling the diffusive behavior of the particles, I was able to generate synthetic data, which served to validate the estimators derived from the statistical model I previously discussed. I believe the approach I have employed offers a more intuitive model of an FCS experiment.

6.4 Making modelling easier with Network Designer

Mathematical models play a critical role in our understanding of biological systems. By creating mathematical representations of biological systems, researchers can gain insights into the underlying mechanisms that govern cellular processes, such as the interactions between DNA, mRNA, and proteins. These models can help to identify key pathways, predict outcomes, and test hypotheses that would otherwise be difficult or impossible to study experimentally. To aid in the creation of complex interaction networks, I developed a new tool called Network Designer. This tool provides an easy graphical interface for modelling and generating the associated equations, enabling rapid prototyping of different mathematical models. This reduces the risk of making small mistakes that are common when writing down large systems of coupled ordinary differential equations. Network Designer was used to model key elements of the circadian system, including the CLOCK:BMAL1 transcription factor binding to DNA in chapter 2 and facilitated protein-protein interactions leading to ternary complex formation in chapter 3. The software is designed to be adaptable, allowing users to incorporate a wide range of models and simulations into their work.

In the future, I aim to expand the capabilities and accessibility of Network Designer so that it can become a valuable tool for researchers to apply mathematical modelling to their systems. To achieve this, I plan to expand the auto-equation feature and library of existing models, as well as allowing users to define models in new and different ways, such as through new natural language AI systems. A benefit of mathematical models is that they can provide a framework for integrating and synthesising data from multiple sources, I envision Network Designer becoming a central repository for defining and curating models of biomolecular systems. This would allow the user to bring together experimental data and models in one place.

6.5 A new multistage model of the circadian clock

While substantial progress has been made in the genetic understanding of the circadian clock, a comprehensive understanding of how the molecular interactions within the network contribute to the generation of stable 24-hour rhythms have been lacking. Recent studies and the work presented in this thesis present new understandings of the clock and mechanistic models which underpin the generation of robust 24-hour rhythms. The circadian clock uses molecules, primarily comprising DNA, RNA, and proteins, to drive daily cycles. The abundance of these molecules as well as the sequence and types of interactions between them are key to defining how these oscillations persist and their properties. In the circadian clock compromises are made between characteristics such as energetic requirements vs robustness, and plasticity vs stability during cross-talk. Stochastic modelling by Gonze and Goldbeter (2006) demonstrated that for simple circadian networks, whilst maintaining the stoichiometric relationship between the molecular components, reducing the overall 'size' of the oscillator impacts robustness of rhythms as well as the ability to effectively couple with other cells. Given that other studies and data within this thesis place circadian molecular components as low abundance, namely 10-100 mRNA and 1,000 - 20,000 proteins per cell (Phillips et al., 2021; Smyllie et al., 2016; Narumi et al., 2016), a key question to ask is *why* are the numbers so low? Whilst further studies are required to fully answer this question at an evolutionary level, we can shed light on *how* the circadian clock can operate in this regime.

Since the first mathematical descriptions by Goodwin, circadian clocks are thought to operate as stable limit cycles that are insensitive to small perturbations and not solely dependent on their initial conditions (Goodwin, 1965; Gonze and Ruoff, 2021). The clock accounts for this through constant entrainment to the external environment starting with photic cues entering the retina to synchronise the Suprachiasmatic Nucleus (SCN), followed by SCN coordinating timing in peripheral tissues via neuroendocrine and other cues (Panda et al., 2002b; van der Horst et al., 1999; Kaneko et al., 1981; Meyer-Bernstein et al., 1999). In tissue explants, free running circadian rhythms generally rapidly desynchronise. An exception is the robust SCN which can oscillate for months ex vivo due to strong inter-cellular coupling (Webb et al., 2009; Welsh et al., 2004). Additionally, non-linearity, delayed repression, and multiple loops have all been shown to be vital in creating a functional intracellular clock mechanism (Kim and Forger, 2012; Heidebrecht et al., 2020).

Emerging models of the clock have demonstrated how the molecular circadian clock operates via multiple complimentary and competitive mechanisms, leading to a model in which the circadian clock operates as an intricate oscillator made up of multiple feedback loops. These mechanisms are thought to operate primarily as a transcription translation feedback loop (TTFL) with negative repression of CLOCK:BMAL1 by PERs and CRYs, as well as additional modulating positive and negative loops comprising Rev-Erbs, RORs, and others (Takahashi, 2017). Evidence shows that all-important PER/CRY repression occurs via three distinct mechanisms: displacement, sequestration, and blocking (Xu et al., 2015; Ye et al., 2011, 2014; Cao et al., 2021; Kim, 2016; Jeong et al., 2022). Specifically, PER proteins work with the CRYs to displace and sequester CLOCK:BMAL1 from DNA, whereas CRY1 is capable of directly binding BMAL1 to block transactivation. Modelling has shown combinations of these mechanisms creates ultra-sensitivity, which rapidly 'switches on' repression over very small changes in the concentrations of the repressor proteins PER and CRY (Jeong et al., 2022). My work supports this model as I have shown that the residency of CLOCK:BMAL1 on E-Box DNA sites scales exponentially with the concentration of PER and CRY (chapter 2). Thus, although transcription can be repressed by any one of these three mechanisms alone, this and other's studies have presented evidence for the circadian clock operating synergistically via all three (Xu et al., 2015; Ye et al., 2011, 2014; Cao et al., 2021; Kim, 2016; Jeong et al., 2022).

Prior to transcriptional repression, protein-protein interactions define when, where, and how many of each of the components can come together to affect their function. Models of the circadian clock have suggested that there is a requirement of one-to-one stoichiometry between repressors PER/CRY and activators CLOCK/BMAL1, which is relaxed by strong binding (< 25 nM) (Kim and Forger, 2012; Heidebrecht et al., 2020). Quantitative data of the concentration of all four of these proteins does not support this strict one-to-one stoichiometric requirement, but is reconciled by data in this thesis of very strong binding between CLOCK-BMAL1 ($38 - 83$ nM), PER-CRY ($20 - 60$ nM), and BMAL1-CRY1 ($15 - 26$ nM) (chapter 3). Strong affinities between these repressive and activating proteins confers ultra-sensitivity to changing levels of repressors, enabling rapid switching from repression to activation. Studies of theoretical oscillatory models have shown that robust cycles are promoted by negative feedback loops with delay primarily made up from 'branching' reactions, i.e. degradation and nuclear export of proteins, rather than 'in-loop' kinetics such as synthesis and phosphorylation (Kurosawa and Iwasa, 2002). In this thesis, BMAL1 was found to mediate CLOCK nuclear entry as well as the PERs and CRYs reciprocally altering each other's sub-cellular localisations. Additionally, PER1 and BMAL1 also up regulated both CRY1 and CRY2 protein expression, which was capable of generating circadian cycles without the need for rhythmic transcriptional control of CRY2. Again, modelling has shown that dimerisation plays a key role in stabilising the limit cycle of the clock as coupling two separate oscillators, here PER and CRY, can generate the required ultra-sensitivity (Kurosawa et al., 2002). Remarkably, interaction measurements suggest both PERs and CRYs homo- and hetero-dimerise strongly, and also readily form three-way complexes with CK1δ (chapter 3). This is the first direct evidence of these homo- and hetero-dimers in mammalian cells. The ternary CK1-PER-CRY complex is essential for displacement type repres-

sion of CLOCK:BMAL1, and represents one of the few abundant multi-meric protein complexes in the circadian clock (Ye et al., 2011, 2014; Cao et al., 2021). The requirement for protein complex formation for effective negative repression buffers against the significant molecular noise within and between cells, and explains how oscillations can be generated even when rhythmic transcription of BMAL1, CRY1, or CRY2 is removed (Liu et al., 2008; McManus et al., 2022) (chapter 3).

This thesis therefore utilises new analytical tools and approaches to provide missing quantitative data on many components and process of the circadian clock. Moreover, modelling of my data reveals how long 24-hour circadian rhythms – much longer than the half-life of the most long-lived oscillatory component – are built up from combinations of series of multiple rapid biochemical processes occurring on the order of seconds to minutes.

While the work contained in this thesis does not address medical translational potential or broader applications, there are major potential benefits from use of the tools and reagents developed in the course of my work. There is much current interest in the potential use of small-molecule activators/inhibitors of the circadian clock in relation to the clinical treatment of circadian disorders, in addition other pathologies. Additionally, a variety of studies have indicated the the timing of drug dosing, particularly in the case of cancer treatments, can lead to improved health outcomes by only exposing tissues to drugs when they will be most able to tolerate toxicity (Lévi et al., 2010; Annabelle Ballesta et al., 2017; Cederroth et al., 2019; Ruben et al., 2019).

The mathematical and biological approaches I have developed here offer an ideal testing platform for the screening of candidate circadian modulating drugs, and will allow direct measurements of how potential small molecule compounds impact on the protein components that lie at the core of the circadian clockwork. These issues are yet to be explored.

References

Abe J, Hiyama TB, Mukaiyama A, Son S, Mori T, Saito S, Osako M, Wolanin J, Yamashita E, Kondo T, Akiyama S. Atomic-scale origins of slowness in the cyanobacterial circadian clock. Science. 2015 Jul; 349(6245):312–316. https://doi.org/10.1126/science.1261040, doi: 10.1126/science.1261040, publisher: American Association for the Advancement of Science.

Ahmad M, Jarillo JA, Cashmore AR. Chimeric Proteins between cry1 and cry2 Arabidopsis Blue Light Photoreceptors Indicate Overlapping Functions and Varying Protein Stability. The Plant Cell. 1998 Feb; 10(2):197–207. https://doi.org/10.1105/tpc.10.2.197, doi: 10.1105/tpc.10.2.197.

Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974 Dec; 19(6):716–723. doi: 10.1109/TAC.1974.1100705.

Alon U. Network motifs: theory and experimental approaches. Nature Reviews Genetics. 2007 Jun; 8(6):450–461. https://doi.org/10.1038/nrg2102, doi: 10.1038/nrg2102.

Alvarenga DO, Fiore MF, Varani AM. A Metagenomic Approach to Cyanobacterial Genomics. Frontiers in Microbiology. 2017; 8. https://www.frontiersin.org/articles/10.3389/fmicb.2017.00809.

Annabelle Ballesta, Pasquale F Innominato, Robert Dallmann, David A Rand, Francis A Lévi. Systems Chronotherapeutics. Pharmacological Reviews. 2017 Apr; 69(2):161. http://pharmrev.aspetjournals.org/content/69/2/161.abstract, doi: 10.1124/pr.116.013441.

Antoch MP, Song EJ, Chang AM, Vitaterna MH, Zhao Y, Wilsbacher LD, Sangoram AM, King DP, Pinto LH, Takahashi JS. Functional Identification of the Mouse Circadian Clock Gene by Transgenic BAC Rescue. Cell. 1997 May; 89(4):655–667. https://doi.org/10.1016/S0092-8674(00)80246-9, doi: 10.1016/S0092-8674(00)80246-9, publisher: Elsevier.

Aschoff J. Exogenous and Endogenous Components in Circadian Rhythms. Cold Spring Harbor Symposia on Quantitative Biology. 1960; 25:11–28. http://symposium.cshlp.org/content/25/11.short, doi: 10.1101/SQB.1960.025.01.004, _eprint: http://symposium.cshlp.org/content/25/11.full.pdf+html.

Axelrod D, Koppel DE, Schlessinger J, Elson E, Webb WW. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. Biophysical Journal. 1976 Sep; 16(9):1055–1069. https://www.sciencedirect.com/science/article/pii/S0006349576857554, doi: 10.1016/S0006-3495(76)85755-4.

Azpeitia E, Wagner A. Short Residence Times of DNA-Bound Transcription Factors Can Reduce Gene Expression Noise and Increase the Transmission of Information in a Gene Regulation System. Frontiers in Molecular Biosciences. 2020; 7. https://www.frontiersin.org/articles/10.3389/fmolb.2020.00067.

Bacia K, Kim SA, Schwille P. Fluorescence cross-correlation spectroscopy in living cells. Nature Methods. 2006 Feb; 3(2):83–89. https://doi.org/10.1038/nmeth822, doi: 10.1038/nmeth822.

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 May; 30(2):525–536. https://doi.org/10.1016/S0896-6273(01)00302-6, doi: 10.1016/S0896-6273(01)00302-6, publisher: Elsevier.

Bagnall J, Leedale J, Taylor SE, Spiller DG, White MRH, Sharkey KJ, Bearon RN, Sée V. Tight Control of Hypoxia-inducible Factor- Transient Dynamics Is Essential for Cell Survival in Hypoxia. Journal of Biological Chemistry. 2014 Feb; 289(9):5549–5564. https://doi.org/10.1074/jbc.M113.500405, doi: 10.1074/jbc.M113.500405, publisher: Elsevier.

Balsalobre A, Marcacci L, Schibler U. Multiple signaling pathways elicit circadian gene expression in cultured Rat-1 fibroblasts. Current Biology. 2000 Oct; 10(20):1291–1294. https://doi.org/10.1016/S0960-9822(00)00758-2, doi: 10.1016/S0960-9822(00)00758-2, publisher: Elsevier.

Bargiello TA, Jackson FR, Young MW. Restoration of circadian behavioural rhythms by gene transfer in Drosophila. Nature. 1984 Dec; 312(5996):752–754. https://doi.org/10.1038/312752a0, doi: 10.1038/312752a0.

Bass J, Lazar MA. Circadian time signatures of fitness and disease. Science. 2016 Nov; 354(6315):994–999. https://doi.org/10.1126/science.aah4965, doi: 10.1126/science.aah4965, publisher: American Association for the Advancement of Science.

Baum K, Politi AZ, Kofahl B, Steuer R, Wolf J. Feedback, Mass Conservation and Reaction Kinetics Impact the Robustness of Cellular Oscillations. PLOS Computational Biology. 2016 Dec; 12(12):e1005298. https://doi.org/10.1371/journal.pcbi.1005298, doi: 10.1371/journal.pcbi.1005298, publisher: Public Library of Science.

Becker-Weimann S, Wolf J, Herzel H, Kramer A. Modeling Feedback Loops of the Mammalian Circadian Oscillator. Biophysical Journal. 2004 Nov; 87(5):3023–3034. https://doi.org/10.1529/biophysj.104.040824, doi: 10.1529/biophysj.104.040824, publisher: Elsevier.

Bertsimas D, Tsitsiklis J. Simulated Annealing. Statistical Science. 1993 Feb; 8(1):10–15. https://doi.org/10.1214/ss/1177011077, doi: 10.1214/ss/1177011077.

Betancourt M, A Conceptual Introduction to Hamiltonian Monte Carlo. arXiv; 2017. https://arxiv.org/abs/1701.02434, doi: 10.48550/ARXIV.1701.02434.

Beytebiere JR, Trott AJ, Greenwell BJ, Osborne CA, Vitet H, Spence J, Yoo SH, Chen Z, Takahashi JS, Ghaffari N, Menet JS. Tissue-specific BMAL1 cistromes reveal that rhythmic transcription is associated with rhythmic enhancer-enhancer interactions. Genes Dev. 2019; 33(5-6):294–309. https://www.ncbi.nlm.nih.gov/pubmed/30804225, doi: 10.1101/gad.322198.118, type: Journal Article.

Bieler J, Cannavo R, Gustafson K, Gobet C, Gatfield D, Naef F. Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. Molecular Systems Biology. 2014 Jul; 10(7):739. https://doi.org/10.15252/msb.20145218, doi: 10.15252/msb.20145218, publisher: John Wiley & Sons, Ltd.

Billingsley P. Probability and measure, 3rd ed. John Wiley & Sons; 2008.

Blinov ML, Schaff JC, Vasilescu D, Moraru II, Bloom JE, Loew LM. Compartmental and Spatial Rule-Based Modeling with Virtual Cell. Biophysical Journal. 2017 Oct; 113(7):1365–1372. https://doi.org/10.1016/j.bpj.2017.08.022, doi: 10.1016/j.bpj.2017.08.022, publisher: Elsevier.

Briggs GE, Haldane JBS. A Note on the Kinetics of Enzyme Action. Biochemical Journal. 1925 Jan; 19(2):338–339. https://doi.org/10.1042/bj0190338, doi: 10.1042/bj0190338.

Buhr ED, Yoo SH, Takahashi JS. Temperature as a Universal Resetting Cue for Mammalian Circadian Oscillators. Science. 2010 Oct; 330(6002):379–385. https://doi.org/10.1126/science.1195262, doi: 10.1126/science.1195262, publisher: American Association for the Advancement of Science.

Buijs RM, la Fleur SE, Wortel J, van Heyningen C, Zuiddam L, Mettenleiter TC, Kalsbeek A, Nagai K, Niijima A. The suprachiasmatic nucleus balances sympathetic and parasympathetic output to peripheral organs through separate preautonomic neurons. Journal of Comparative Neurology. 2003 Sep; 464(1):36–48. https://doi.org/10.1002/cne.10765, doi: 10.1002/cne.10765, publisher: John Wiley & Sons, Ltd.

Bunger MK, Wilsbacher LD, Moran SM, Clendenin C, Radcliffe LA, Hogenesch JB, Simon MC, Takahashi JS, Bradfield CA. Mop3 Is an Essential Component of the Master Circadian Pacemaker in Mammals. Cell. 2000 Dec; 103(7):1009–1017. https://doi.org/10.1016/S0092-8674(00)00205-1, doi: 10.1016/S0092-8674(00)00205-1, publisher: Elsevier.

Burton J, Manning CS, Rattray M, Papalopulu N, Kursawe J. Inferring kinetic parameters of oscillatory gene regulation from single cell time-series data. Journal of The Royal Society Interface. 2021; 18(182):20210393. https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2021.0393, doi: 10.1098/rsif.2021.0393, _eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2021.0393.

Calderazzo S, Brancaccio M, Finkenstädt B. Filtering and inference for stochastic oscillators with distributed delays. Bioinformatics. 2019 Apr; 35(8):1380–1387. https://doi.org/10.1093/bioinformatics/bty782, doi: 10.1093/bioinformatics/bty782.

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). https://www.ncbi.nlm.nih.gov/pubmed/33443219, doi: 10.1073/pnas.2021174118, type: Journal Article.

Carrero G, McDonald D, Crawford E, de Vries G, Hendzel MJ. Using FRAP and mathematical modeling to determine the in vivo kinetics of nuclear proteins. Methods. 2003 Jan; 29(1):14–28. https://www.sciencedirect.com/science/article/pii/S1046202302002888, doi: 10.1016/S1046-2023(02)00288-8.

Cederroth CR, Albrecht U, Bass J, Brown SA, Dyhrfjeld-Johnsen J, Gachon F, Green CB, Hastings MH, Helfrich-Förster C, Hogenesch JB, Lévi F, Loudon A, Lundkvist GB, Meijer JH, Rosbash M, Takahashi JS, Young M, Canlon B. Medicine in the Fourth Dimension. Cell Metabolism. 2019 Aug; 30(2):238–250. https://doi.org/10.1016/j.cmet.2019.06.019, doi: 10.1016/j.cmet.2019.06.019, publisher: Elsevier.

Chakrabarti A, Ghosh JK. AIC, BIC and Recent Advances in Model Selection. In: Bandyopadhyay PS, Forster MR, editors. *Philosophy of Statistics*, vol. 7 Amsterdam: North-Holland; 2011.p. 583–605. https://www.sciencedirect.com/science/article/pii/B9780444518620500186, doi: 10.1016/B978-0-444-51862-0.50018-6.

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green Fluorescent Protein as a Marker for Gene Expression. Science. 1994 Feb; 263(5148):802–805. https://doi.org/10.1126/science.8303295, doi: 10.1126/science.8303295, publisher: American Association for the Advancement of Science.

Chen Y, Müller JD, So PTC, Gratton E. The Photon Counting Histogram in Fluorescence Fluctuation Spectroscopy. Biophysical Journal. 1999 Jul; 77(1):553–567. https://www.sciencedirect.com/science/article/pii/S0006349599769122, doi: 10.1016/S0006-3495(99)76912-2.

Chiou YY, Yang Y, Rashid N, Ye R, Selby CP, Sancar A. Mammalian Period represses and de-represses transcription by displacing CLOCK–BMAL1 from promoters in a Cryptochrome-dependent manner. Proceedings of the National Academy of Sciences. 2016 Oct; 113(41):E6072–E6079. https://doi.org/10.1073/pnas.1612917113, doi: 10.1073/pnas.1612917113, publisher: Proceedings of the National Academy of Sciences.

Chong KH, Samarasinghe S, Kulasiri D. Mathematical modelling of p53 basal dynamics and DNA damage response. Mathematical Biosciences. 2015 Jan; 259:27–42. https://www.sciencedirect.com/science/article/pii/S0025556414002259, doi: 10.1016/j.mbs.2014.10.010.

Costa MJ, Finkenstädt B, Roche V, Lévi F, Gould PD, Foreman J, Halliday K, Hall A, Rand DA. Inference on periodicity of circadian time series. Biostatistics. 2013 Sep; 14(4):792–806. https://doi.org/10.1093/biostatistics/kxt020, doi: 10.1093/biostatistics/kxt020.

Darwin C 1809-1882. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London : John Murray, 1859; 1859. https://search.library.wisc.edu/catalog/9934839413602122.

Davis AW. On the theory of birth, death and diffusion processes. Journal of Applied Probability. 1965; 2(2):293–322. https://www.cambridge.org/core/article/on-the-theory-of-birth-death-and-diffusion-processes/CDEBAA22A8E7AAC2BA1D75877F2814F7, doi: 10.2307/3212196, edition: 2016/07/14 Publisher: Cambridge University Press.

DeBruyne JP, Weaver DR, Reppert SM. CLOCK and NPAS2 have overlapping roles in the suprachiasmatic circadian clock. Nature Neuroscience. 2007 May; 10(5):543–545. https://doi.org/10.1038/nn1884, doi: 10.1038/nn1884.

Du C, Kou SC. Statistical Methodology in Single-Molecule Experiments. Statistical Science. 2020 Feb; 35(1):75–91. https://doi.org/10.1214/19-STS752, doi: 10.1214/19-STS752.

Dyar KA, Hubert MJ, Mir AA, Ciciliot S, Lutter D, Greulich F, Quagliarini F, Kleinert M, Fischer K, Eichmann TO, Wright LE, Peña Paz MI, Casarin A, Pertegato V, Romanello V, Albiero M, Mazzucco S, Rizzuto R, Salviati L, Biolo G, et al. Transcriptional programming of lipid and amino acid metabolism by the skeletal muscle circadian clock. PLOS Biology. 2018 Aug; 16(8):e2005886. https://doi.org/10.1371/journal.pbio.2005886, doi: 10.1371/journal.pbio.2005886, publisher: Public Library of Science.

Eide EJ, Woolf MF, Kang H, Woolf P, Hurst W, Camacho F, Vielhaber EL, Giovanni A, Virshup DM. Control of Mammalian Circadian Rhythm by CKI -Regulated Proteasome-Mediated PER2 Degradation. Molecular and Cellular Biology. 2005 Apr; 25(7):2795–2807. https://doi.org/10.1128/MCB.25.7.2795-2807.2005, doi: 10.1128/MCB.25.7.2795-2807.2005, publisher: American Society for Microbiology.

Einstein A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Annalen der Physik. 1905 Jan; 322(8):549–560. https://doi.org/10.1002/andp.19053220806, doi: 10.1002/andp.19053220806, publisher: John Wiley & Sons, Ltd.

Erica V, Erik E, Ann R, Zhong-Hua G, Virshup DM. Nuclear Entry of the Circadian Regulator mPER1 Is Controlled by Mammalian Casein Kinase I . Molecular and Cellular Biology. 2000 Jul; 20(13):4888–4899. https://doi.org/10.1128/MCB.20.13.4888-4899.2000, doi: 10.1128/MCB.20.13.4888-4899.2000, publisher: American Society for Microbiology.

Featherstone K, Hey K, Momiji H, McNamara AV, Patist AL, Woodburn J, Spiller DG, Christian HC, McNeilly AS, Mullins JJ, Finkenstädt BF, Rand DA, White MR, Davis JR. Spatially coordinated dynamic gene transcription in living pituitary tissue. eLife. 2016 Feb; 5:e08494. https://doi.org/10.7554/eLife.08494, doi: 10.7554/eLife.08494, publisher: eLife Sciences Publications, Ltd.

Fick A. Ueber Diffusion. Annalen der Physik. 1855 Jan; 170(1):59–86. https://doi.org/10.1002/andp.18551700105, doi: 10.1002/andp.18551700105, publisher: John Wiley & Sons, Ltd.

Finney A, Hucka M. Systems biology markup language: Level 2 and beyond. Biochemical Society Transactions. 2003 Dec; 31(6):1472–1473. https://doi.org/10.1042/bst0311472, doi: 10.1042/bst0311472, _eprint: https://portlandpress.com/biochemsoctrans/article-pdf/31/6/1472/537721/bst0311472.pdf.

Fisher RA. Tests of significance in harmonic analysis. Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character. 1929 Aug; 125(796):54–59. https://doi.org/10.1098/rspa.1929.0151, doi: 10.1098/rspa.1929.0151, publisher: Royal Society.

Forger DB. Biological clocks, rhythms, and oscillations: the theory of biological timekeeping. . 2017; Publisher: MIT Press.

Forger DB, Peskin CS. A detailed predictive model of the mammalian circadian clock. Proceedings of the National Academy of Sciences. 2003 Dec; 100(25):14806–14811. https://doi.org/10.1073/pnas.2036281100, doi: 10.1073/pnas.2036281100, publisher: Proceedings of the National Academy of Sciences.

Forger DB, Peskin CS. Stochastic simulation of the mammalian circadian clock. Proceedings of the National Academy of Sciences. 2005 Jan; 102(2):321–324. https://doi.org/10.1073/pnas.0408465102, doi: 10.1073/pnas.0408465102, publisher: Proceedings of the National Academy of Sciences.

Freedman MS, Lucas RJ, Soni B, von Schantz M, Muñoz M, David-Gray Z, Foster R. Regulation of Mammalian Circadian

Behavior by Non-rod, Non-cone, Ocular Photoreceptors. Science. 1999 Apr; 284(5413):502–504. https://doi.org/10.1126/science.284.5413.502, doi: 10.1126/science.284.5413.502, publisher: American Association for the Advancement of Science.

Fribourgh JL, Srivastava A, Sandate CR, Michael AK, Hsu PL, Rakers C, Nguyen LT, Torgrimson MR, Parico GCG, Tripathi S, Zheng N, Lander GC, Hirota T, Tama F, Partch CL. Dynamics at the serine loop underlie differential affinity of cryptochromes for CLOCK:BMAL1 to control circadian timing. Elife. 2020; 9. https://www.ncbi.nlm.nih.gov/pubmed/32101164, doi: 10.7554/eLife.55275, type: Journal Article.

Froy O, Gotter AL, Casselman AL, Reppert SM. Illuminating the circadian clock in monarch butterfly migration. Science. 2003 May; 300(5623):1303–1305. http://europepmc.org/abstract/MED/12764200, doi: 10.1126/science.1084874.

Fussenegger M, Bailey JE, Varner J. A mathematical model of caspase function in apoptosis. Nature Biotechnology. 2000 Jul; 18(7):768–774. https://doi.org/10.1038/77589, doi: 10.1038/77589.

Fustin JM, O'Neill JS, Hastings MH, Hazlerigg DG, Dardente H. Cry1 Circadian Phase in vitro: Wrapped Up with an E-Box. Journal of Biological Rhythms. 2009 Feb; 24(1):16–24. https://doi.org/10.1177/0748730408329267, doi: 10.1177/0748730408329267, publisher: SAGE Publications Inc.

Gan S, O'Shea EK. An Unstable Singularity Underlies Stochastic Phasing of the Circadian Clock in Individual Cyanobacterial Cells. Molecular Cell. 2017 Aug; 67(4):659–672.e12. https://www.sciencedirect.com/science/article/pii/S1097276517305099, doi: 10.1016/j.molcel.2017.07.015.

Gatfield D, Schibler U. Proteasomes Keep the Circadian Clock Ticking. Science. 2007 May; 316(5828):1135–1136. https://doi.org/10.1126/science.1144165, doi: 10.1126/science.1144165, publisher: American Association for the Advancement of Science.

Gekakis N, Saez L, Delahaye-Brown AM, Myers MP, Sehgal A, Young MW, Weitz CJ. Isolation of timeless by PER Protein Interaction: Defective Interaction Between timeless Protein and Long-Period Mutant PERL. Science. 1995 Nov; 270(5237):811–815. https://doi.org/10.1126/science.270.5237.811, doi: 10.1126/science.270.5237.811, publisher: American Association for the Advancement of Science.

Gekakis N, Staknis D, Nguyen HB, Davis FC, Wilsbacher LD, King DP, Takahashi JS, Weitz CJ. Role of the CLOCK Protein in the Mammalian Circadian Mechanism. Science. 1998 Jun; 280(5369):1564–1569. https://doi.org/10.1126/science.280.5369.1564, doi: 10.1126/science.280.5369.1564, publisher: American Association for the Advancement of Science.

Gibbs JE, Blaikley J, Beesley S, Matthews L, Simpson KD, Boyce SH, Farrow SN, Else KJ, Singh D, Ray DW, Loudon ASI. The nuclear receptor REV-ERB mediates circadian regulation of innate immunity through selective regulation of inflammatory cytokines. Proceedings of the National Academy of Sciences. 2012 Jan; 109(2):582–587. https://doi.org/10.1073/pnas.1106750109, doi: 10.1073/pnas.1106750109, publisher: Proceedings of the National Academy of Sciences.

Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics. 1976; 22(4):403–434. doi: 10.1016/0021-9991(76)90041-3, type: Journal Article.

Gillespie DT. The chemical Langevin equation. The Journal of Chemical Physics. 2000 Jul; 113(1):297–306. https://doi.org/10.1063/1.481811, doi: 10.1063/1.481811, publisher: American Institute of Physics.

Goldbeter A. A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase. Proceedings of the National Academy of Sciences. 1991 Oct; 88(20):9107–9111. https://doi.org/10.1073/pnas.88.20.9107, doi: 10.1073/pnas.88.20.9107, publisher: Proceedings of the National Academy of Sciences.

Goldbeter A. Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour. Cambridge: Cambridge University Press; 1996. https://www.cambridge.org/core/books/biochemical-oscillations-and-cellular-rhythms/F47CCE61995A40AE650A527B4395D200, doi: 10.1017/CBO9780511608193.

Gonze D, Goldbeter A. Circadian rhythms and molecular noise. Chaos. 2006; 16(2):026110. https://www.ncbi.nlm.nih.gov/pubmed/16822042, doi: 10.1063/1.2211767, type: Journal Article.

Gonze D, Bernard S, Waltermann C, Kramer A, Herzel H. Spontaneous Synchronization of Coupled Circadian Oscillators. Biophysical Journal. 2005 Jul; 89(1):120–129. https://doi.org/10.1529/biophysj.104.058388, doi: 10.1529/biophysj.104.058388, publisher: Elsevier.

Gonze D, Ruoff P. The Goodwin Oscillator and its Legacy. Acta Biotheoretica. 2021 Dec; 69(4):857–874. https://doi.org/10.1007/s10441-020-09379-8, doi: 10.1007/s10441-020-09379-8.

Goodwin BC. Oscillatory behavior in enzymatic control processes. Advances in Enzyme Regulation. 1965 Jan; 3:425–437. https://www.sciencedirect.com/science/article/pii/0065257165900671, doi: 10.1016/0065-2571(65)90067-1.

Gotter AL, Manganaro T, Weaver DR, Kolakowski LF, Possidente B, Sriram S, MacLaughlin DT, Reppert SM. A timeless function for mouse Timeless. Nature Neuroscience. 2000 Aug; 3(8):755–756. https://doi.org/10.1038/77653, doi: 10.1038/77653.

Green CB, Takahashi JS, Bass J. The Meter of Metabolism. Cell. 2008 Sep; 134(5):728–742. https://doi.org/10.1016/j.cell.2008.08.022, doi: 10.1016/j.cell.2008.08.022, publisher: Elsevier.

Griffith JS. Mathematics of cellular control processes I. Negative feedback to one gene. Journal of Theoretical Biol-

ogy. 1968 Aug; 20(2):202–208. https://www.sciencedirect.com/science/article/pii/0022519368901896, doi: 10.1016/0022-5193(68)90189-6.

Gunawardena J. Multisite protein phosphorylation makes a good threshold but can be a poor switch. Proceedings of the National Academy of Sciences. 2005 Oct; 102(41):14617–14622. https://doi.org/10.1073/pnas.0507322102, doi: 10.1073/pnas.0507322102, publisher: Proceedings of the National Academy of Sciences.

Gustafson CL, Parsley NC, Asimgil H, Lee HW, Ahlbach C, Michael AK, Xu H, Williams OL, Davis TL, Liu AC, Partch CL. A Slow Conformational Switch in the BMAL1 Transactivation Domain Modulates Circadian Rhythms. Mol Cell. 2017; 66(4):447–457 e7. https://www.ncbi.nlm.nih.gov/pubmed/28506462, doi: 10.1016/j.molcel.2017.04.011, type: Journal Article.

Gustafson CL, Partch CL. Emerging Models for the Molecular Basis of Mammalian Circadian Timing. Biochemistry. 2015 Jan; 54(2):134–149. https://doi.org/10.1021/bi500731f, doi: 10.1021/bi500731f, publisher: American Chemical Society.

Gómez-Schiavon M, Chen LF, West AE, Buchler NE. BayFish: Bayesian inference of transcription dynamics from population snapshots of single-molecule RNA FISH in single cells. Genome Biology. 2017 Sep; 18(1):164. https://doi.org/10.1186/s13059-017-1297-9, doi: 10.1186/s13059-017-1297-9.

Hardin PE, Hall JC, Rosbash M. Feedback of the Drosophila period gene product on circadian cycling of its messenger RNA levels. Nature. 1990 Feb; 343(6258):536–540. https://doi.org/10.1038/343536a0, doi: 10.1038/343536a0.

Harris CR, Millman KJ, Walt SJvd, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, Kerkwijk MHv, Brett M, Haldane A, Río JFd, Wiebe M, Peterson P, Gérard-Marchant P, et al. Array programming with NumPy. Nature. 2020 Sep; 585(7825):357–362. https://doi.org/10.1038/s41586-020-2649-2, doi: 10.1038/s41586-020-2649-2, publisher: Springer Science and Business Media LLC.

Hastings MH, Maywood ES, Brancaccio M. Generation of circadian rhythms in the suprachiasmatic nucleus. Nature Reviews Neuroscience. 2018 Aug; 19(8):453–469. https://doi.org/10.1038/s41583-018-0026-z, doi: 10.1038/s41583-018-0026-z.

Hastings MH, Reddy AB, Maywood ES. A clockwork web: circadian timing in brain and periphery, in health and disease. Nature Reviews Neuroscience. 2003 Aug; 4(8):649–661. https://doi.org/10.1038/nrn1177, doi: 10.1038/nrn1177.

Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970 Apr; 57(1):97–109. https://doi.org/10.1093/biomet/57.1.97, doi: 10.1093/biomet/57.1.97.

Hattar S, Liao HW, Takao M, Berson DM, Yau KW. Melanopsin-Containing Retinal Ganglion Cells: Architecture, Projections, and Intrinsic Photosensitivity. Science. 2002 Feb; 295(5557):1065–1070. https://doi.org/10.1126/science.1069609, doi: 10.1126/science.1069609, publisher: American Association for the Advancement of Science.

Haupts U, Maiti S, Schwille P, Webb WW. Dynamics of fluorescence fluctuations in green fluorescent protein observed by fluorescence correlation spectroscopy. Proceedings of the National Academy of Sciences. 1998; 95(23):13573–13578. https://www.pnas.org/doi/abs/10.1073/pnas.95.23.13573, doi: 10.1073/pnas.95.23.13573, _eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.95.23.13573.

Haynes W. Maximum Likelihood Estimation. In: Dubitzky W, Wolkenhauer O, Cho KH, Yokota H, editors. *Encyclopedia of Systems Biology* New York, NY: Springer New York; 2013.p. 1190–1191. https://doi.org/10.1007/978-1-4419-9863-7_1235, doi: 10.1007/978-1-4419-9863-7_1235.

Heidebrecht B, Chen J, Tyson JJ. Mathematical Analysis of Robustness of Oscillations in Models of the Mammalian Circadian Clock. bioRxiv. 2020 Jan; p. 2020.09.15.297648. http://biorxiv.org/content/early/2020/09/15/2020.09.15.297648.abstract, doi: 10.1101/2020.09.15.297648.

Hettich J, Gebhardt JCM. Transcription factor target site search and gene regulation in a background of unspecific binding sites. Journal of Theoretical Biology. 2018 Oct; 454:91–101. https://www.sciencedirect.com/science/article/pii/S0022519318302844, doi: 10.1016/j.jtbi.2018.05.037.

Hey KL, Momiji H, Featherstone K, Davis JRE, White MRH, Rand DA, Finkenstädt B. A stochastic transcriptional switch model for single cell imaging data. Biostatistics. 2015 Oct; 16(4):655–669. https://doi.org/10.1093/biostatistics/kxv010, doi: 10.1093/biostatistics/kxv010.

Hogenesch JB, Chan WK, Jackiw VH, Brown RC, Gu YZ, Pray-Grant M, Perdew GH, Bradfield CA. Characterization of a Subset of the Basic-Helix-Loop-Helix-PAS Superfamily That Interacts with Components of the Dioxin Signaling Pathway *. Journal of Biological Chemistry. 1997 Mar; 272(13):8581–8593. https://doi.org/10.1074/jbc.272.13.8581, doi: 10.1074/jbc.272.13.8581, publisher: Elsevier.

van der Horst GTJ, Muijtjens M, Kobayashi K, Takano R, Kanno Si, Takao M, de Wit J, Verkerk A, Eker APM, Leenen Dv, Buijs R, Bootsma D, Hoeijmakers JHJ, Yasui A. Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. Nature. 1999 Apr; 398(6728):627–630. https://doi.org/10.1038/19323, doi: 10.1038/19323.

Huang N, Chelliah Y, Shan Y, Taylor CA, Yoo SH, Partch C, Green CB, Zhang H, Takahashi JS. Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. Science. 2012; 337(6091):189–94. https://www.ncbi.nlm.nih.gov/pubmed/22653727, doi: 10.1126/science.1222804, type: Journal Article.

Huang Y, Mayer C, Walch OJ, Bowman C, Sen S, Goldstein C, Tyler J, Forger DB. Distinct Circadian Assessments From Wearable Data Reveal Social Distancing Promoted Internal Desynchrony Between Circadian Markers. Frontiers in Digital Health. 2021; 3. https://www.frontiersin.org/articles/10.3389/fdgth.2021.727504.

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 2003 Mar; 19(4):524–531. https://doi.org/10.1093/bioinformatics/btg015, doi: 10.1093/bioinformatics/btg015, _eprint: https://academic.oup.com/bioinformatics/article-pdf/19/4/524/48903880/bioinformatics_19_4_524.pdf.

Hughes ME, Hogenesch JB, Kornacker K. JTK_CYCLE: An Efficient Nonparametric Algorithm for Detecting Rhythmic Components in Genome-Scale Data Sets. Journal of Biological Rhythms. 2010 Oct; 25(5):372–380. https://doi.org/10.1177/0748730410379711, doi: 10.1177/0748730410379711, publisher: SAGE Publications Inc.

Hurvich CM, Tsai CL. Regression and time series model selection in small samples. Biometrika. 1989 Jun; 76(2):297–307. https://doi.org/10.1093/biomet/76.2.297, doi: 10.1093/biomet/76.2.297.

Höfling F, Franosch T. Anomalous transport in the crowded world of biological cells. Reports on Progress in Physics. 2013 Mar; 76(4):046602. http://dx.doi.org/10.1088/0034-4885/76/4/046602, doi: 10.1088/0034-4885/76/4/046602, publisher: IOP Publishing.

Ikeda M, Nomura M. cDNA Cloning and Tissue-Specific Expression of a Novel Basic Helix–Loop–Helix/PAS Protein (BMAL1) and Identification of Alternatively Spliced Variants with Alternative Translation Initiation Site Usage. Biochemical and Biophysical Research Communications. 1997 Apr; 233(1):258–264. https://www.sciencedirect.com/science/article/pii/S0006291X97963713, doi: 10.1006/bbrc.1997.6371.

Ikeda M, Yu W, Hirai M, Ebisawa T, Honma S, Yoshimura K, Honma KI, Nomura M. cDNA Cloning of a Novel bHLH-PAS Transcription Factor Superfamily Gene, BMAL2: Its mRNA Expression, Subcellular Distribution, and Chromosomal Localization. Biochemical and Biophysical Research Communications. 2000 Aug; 275(2):493–502. https://www.sciencedirect.com/science/article/pii/S0006291X00932480, doi: 10.1006/bbrc.2000.3248.

Inc PT, Collaborative data science; 2015. https://plot.ly, place: Montreal, QC Publisher: Plotly Technologies Inc.

Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology. 1961 Jun; 3(3):318–356. https://www.sciencedirect.com/science/article/pii/S0022283661800727, doi: 10.1016/S0022-2836(61)80072-7.

Jaeger C, Tischkau SA. Role of Aryl Hydrocarbon Receptor in Circadian Clock Disruption and Metabolic Dysfunction. Environ Health Insights. 2016; 10:133–41. https://www.ncbi.nlm.nih.gov/pubmed/27559298, doi: 10.4137/EHI.S38343, type: Journal Article.

Jazani S, Sgouralis I, Shafraz OM, Levitus M, Sivasankar S, Pressé S. An alternative framework for fluorescence correlation spectroscopy. Nature Communications. 2019 Aug; 10(1):3662. https://doi.org/10.1038/s41467-019-11574-2, doi: 10.1038/s41467-019-11574-2.

Jenkins DJ, Finkenstädt B, Rand DA. A temporal switch model for estimating transcriptional activity in gene expression. Bioinformatics. 2013 May; 29(9):1158–1165. https://doi.org/10.1093/bioinformatics/btt111, doi: 10.1093/bioinformatics/btt111.

Jeong EM, Song YM, Kim JK. Combined multiple transcriptional repression mechanisms generate ultrasensitivity and oscillations. Interface Focus. 2022 Apr; 12(3):20210084. https://doi.org/10.1098/rsfs.2021.0084, doi: 10.1098/rsfs.2021.0084, publisher: Royal Society.

Johnson CH, Golden SS, Ishiura M, Kondo T. Circadian clocks in prokaryotes. Molecular Microbiology. 1996 Jul; 21(1):5–11. https://doi.org/10.1046/j.1365-2958.1996.00613.x, doi: 10.1046/j.1365-2958.1996.00613.x, publisher: John Wiley & Sons, Ltd.

Johnson KA, Goody RS. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. Biochemistry. 2011 Oct; 50(39):8264–8269. https://doi.org/10.1021/bi201284u, doi: 10.1021/bi201284u, publisher: American Chemical Society.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug; 596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2, doi: 10.1038/s41586-021-03819-2.

Kaneko M, Kaneko K, Shinsako J, Dallman MF. Adrenal Sensitivity to Adrenocorticotropin Varies Diurnally*. Endocrinology. 1981 Jul; 109(1):70–75. https://doi.org/10.1210/endo-109-1-70, doi: 10.1210/endo-109-1-70.

Karatsoreos IN, Silver R. Chapter 27 - Body Clocks in Health and Disease. In: Conn PM, editor. *Conn's Translational Neuroscience* San Diego: Academic Press; 2017.p. 599–615. https://www.sciencedirect.com/science/article/pii/B9780128023815000439, doi: 10.1016/B978-0-12-802381-5.00043-9.

Keesler GA, Camacho F, Guo Y, Virshup D, Mondadori C, Yao Z. Phosphorylation and destabilization of human period I

clock protein by human casein kinase I . NeuroReport. 2000; 11(5). https://journals.lww.com/neuroreport/Fulltext/2000/04070/Phosphorylation_and_destabilization_of_human.11.aspx.

Keller AD. Model genetic circuits encoding autoregulatory transcription factors. Journal of Theoretical Biology. 1995 Jan; 172(2):169–185. https://www.sciencedirect.com/science/article/pii/S0022519385700148, doi: 10.1006/jtbi.1995.0014.

Kim JK. Protein sequestration versus Hill-type repression in circadian clock models. IET Systems Biology. 2016 Aug; 10(4):125–135. https://doi.org/10.1049/iet-syb.2015.0090, doi: 10.1049/iet-syb.2015.0090, publisher: The Institution of Engineering and Technology.

Kim JK, Forger DB. A mechanism for robust circadian timekeeping via stoichiometric balance. Molecular Systems Biology. 2012 Jan; 8(1):630. https://doi.org/10.1038/msb.2012.62, doi: 10.1038/msb.2012.62, publisher: John Wiley & Sons, Ltd.

Kitchen GB, Cunningham PS, Poolman TM, Iqbal M, Maidstone R, Baxter M, Bagnall J, Begley N, Saer B, Hussell T, Matthews LC, Dockrell DH, Durrington HJ, Gibbs JE, Blaikley JF, Loudon AS, Ray DW. The clock gene Bmal1 inhibits macrophage motility, phagocytosis, and impairs defense against pneumonia. Proceedings of the National Academy of Sciences. 2020 Jan; 117(3):1543–1551. https://doi.org/10.1073/pnas.1915932117, doi: 10.1073/pnas.1915932117, publisher: Proceedings of the National Academy of Sciences.

Klamt S, Saez-Rodriguez J, Gilles ED. Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC Systems Biology. 2007 Jan; 1(1):2. https://doi.org/10.1186/1752-0509-1-2, doi: 10.1186/1752-0509-1-2.

Kloss B, Price JL, Saez L, Blau J, Rothenfluh A, Wesley CS, Young MW. The Drosophila Clock Gene double-time Encodes a Protein Closely Related to Human Casein Kinase I . Cell. 1998 Jul; 94(1):97–107. https://www.sciencedirect.com/science/article/pii/S0092867400812258, doi: 10.1016/S0092-8674(00)81225-8.

Kneen M, Farinas J, Li Y, Verkman AS. Green Fluorescent Protein as a Noninvasive Intracellular pH Indicator. Biophysical Journal. 1998 Mar; 74(3):1591–1599. https://doi.org/10.1016/S0006-3495(98)77870-1, doi: 10.1016/S0006-3495(98)77870-1, publisher: Elsevier.

Koch AA, Bagnall JS, Smyllie NJ, Begley N, Adamson AD, Fribourgh JL, Spiller DG, Meng QJ, Partch CL, Strimmer K, House TA, Hastings MH, Loudon AS. Quantification of protein abundance and interaction defines a mechanism for operation of the circadian clock. eLife. 2022 Mar; 11:e73976. https://doi.org/10.7554/eLife.73976, doi: 10.7554/eLife.73976, publisher: eLife Sciences Publications, Ltd.

Koenker R, Hallock KF. Quantile Regression. Journal of Economic Perspectives. 2001; 15(4):143–156. https://www.aeaweb.org/articles?id=10.1257/jep.15.4.143, doi: 10.1257/jep.15.4.143.

Koike N, Yoo SH, Huang HC, Kumar V, Lee C, Kim TK, Takahashi JS. Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals. Science. 2012; 338(6105):349–354. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3694775, doi: 10.1126/science.1226339, type: Journal Article.

Komin N, Murza AC, Hernández-García E, Toral R. Synchronization and entrainment of coupled circadian oscillators. Interface Focus. 2011 Feb; 1(1):167–176. https://doi.org/10.1098/rsfs.2010.0327, doi: 10.1098/rsfs.2010.0327, publisher: Royal Society.

Konopka RJ, Benzer S. Clock Mutants of Drosophila melanogaster. Proceedings of the National Academy of Sciences. 1971 Sep; 68(9):2112–2116. https://doi.org/10.1073/pnas.68.9.2112, doi: 10.1073/pnas.68.9.2112, publisher: Proceedings of the National Academy of Sciences.

Koppel DE, Axelrod D, Schlessinger J, Elson EL, Webb WW. Dynamics of fluorescence marker concentration as a probe of mobility. Biophysical Journal. 1976 Nov; 16(11):1315–1329. https://www.sciencedirect.com/science/article/pii/S0006349576857761, doi: 10.1016/S0006-3495(76)85776-1.

Kreitzman L, Foster R. The rhythms of life: The biological clocks that control the daily lives of every living thing. Profile books; 2011.

Kurosawa G, Iwasa Y. Saturation of Enzyme Kinetics in Circadian Clock Models. Journal of Biological Rhythms. 2002 Dec; 17(6):568–577. https://doi.org/10.1177/0748730402238239, doi: 10.1177/0748730402238239, publisher: SAGE Publications Inc.

Kurosawa G, Mochizuki A, Iwasa Y. Comparative Study of Circadian Clock Models, in Search of Processes Promoting Oscillation. Journal of Theoretical Biology. 2002 May; 216(2):193–208. https://www.sciencedirect.com/science/article/pii/S0022519302925464, doi: 10.1006/jtbi.2002.2546.

Lambert TJ. FPbase: a community-editable fluorescent protein database. Nature Methods. 2019 Apr; 16(4):277–278. https://doi.org/10.1038/s41592-019-0352-8, doi: 10.1038/s41592-019-0352-8.

Lamia KA, Sachdeva UM, DiTacchio L, Williams EC, Alvarez JG, Egan DF, Vasquez DS, Juguilon H, Panda S, Shaw RJ, Thompson CB, Evans RM. AMPK Regulates the Circadian Clock by Cryptochrome Phosphorylation and Degradation. Science. 2009 Oct; 326(5951):437–440. https://doi.org/10.1126/science.1172156, doi: 10.1126/science.1172156, publisher: American Association for the Advancement of Science.

Lamia KA, Storch KF, Weitz CJ. Physiological significance of a peripheral tissue circadian clock. Proceedings of the Na-

tional Academy of Sciences. 2008 Sep; 105(39):15172–15177. https://doi.org/10.1073/pnas.0806717105, doi: 10.1073/pnas.0806717105, publisher: Proceedings of the National Academy of Sciences.

Lee C, Etchegaray JP, Cagampang FRA, Loudon ASI, Reppert SM. Posttranslational Mechanisms Regulate the Mammalian Circadian Clock. Cell. 2001 Dec; 107(7):855–867. https://www.sciencedirect.com/science/article/pii/S0092867401006109, doi: 10.1016/S0092-8674(01)00610-9.

Lehman M, Silver R, Gladstone W, Kahn R, Gibson M, Bittman E. Circadian rhythmicity restored by neural transplant. Immunocytochemical characterization of the graft and its integration with the host brain. The Journal of Neuroscience. 1987 Jun; 7(6):1626. http://www.jneurosci.org/content/7/6/1626.abstract, doi: 10.1523/JNEUROSCI.07-06-01626.1987.

Leise TL, Wang CW, Gitis PJ, Welsh DK. Persistent Cell-Autonomous Circadian Oscillations in Fibroblasts Revealed by Six-Week Single-Cell Imaging of PER2::LUC Bioluminescence. PLOS ONE. 2012 Mar; 7(3):e33334. https://doi.org/10.1371/journal.pone.0033334, doi: 10.1371/journal.pone.0033334, publisher: Public Library of Science.

Leloup JC, Goldbeter A. Toward a detailed computational model for the mammalian circadian clock. Proceedings of the National Academy of Sciences. 2003 Jun; 100(12):7051–7056. https://doi.org/10.1073/pnas.1132112100, doi: 10.1073/pnas.1132112100, publisher: Proceedings of the National Academy of Sciences.

Leypunskiy E, Lin J, Yoo H, Lee U, Dinner AR, Rust MJ. The cyanobacterial circadian clock follows midday in vivo and in vitro. eLife. 2017 Jul; 6:e23539. https://doi.org/10.7554/eLife.23539, doi: 10.7554/eLife.23539, publisher: eLife Sciences Publications, Ltd.

Li Y, Shan Y, Desai RV, Cox KH, Weinberger LS, Takahashi JS. Noise-driven cellular heterogeneity in circadian periodicity. Proceedings of the National Academy of Sciences. 2020 May; 117(19):10350–10356. https://doi.org/10.1073/pnas.1922388117, doi: 10.1073/pnas.1922388117, publisher: Proceedings of the National Academy of Sciences.

Lightman S, Sassone-Corsi P, Christen Y. Rhythms within rhythms: the importance of oscillations for glucocorticoid hormones. Rhythms within rhythms: the importance of oscillations for glucocorticoid hormones. 2016; Publisher: Springer.

Liu AC, Tran HG, Zhang EE, Priest AA, Welsh DK, Kay SA. Redundant Function of REV-ERB and and Non-Essential Role for Bmal1 Cycling in Transcriptional Regulation of Intracellular Circadian Rhythms. PLOS Genetics. 2008 Feb; 4(2):e1000023. https://doi.org/10.1371/journal.pgen.1000023, doi: 10.1371/journal.pgen.1000023, publisher: Public Library of Science.

Liu AC, Welsh DK, Ko CH, Tran HG, Zhang EE, Priest AA, Buhr ED, Singer O, Meeker K, Verma IM, Doyle FJ III, Takahashi JS, Kay SA. Intercellular Coupling Confers Robustness against Mutations in the SCN Circadian Clock Network. Cell. 2007 May; 129(3):605–616. https://doi.org/10.1016/j.cell.2007.02.047, doi: 10.1016/j.cell.2007.02.047, publisher: Elsevier.

Liu Z, Lavis L, Betzig E. Imaging Live-Cell Dynamics and Structure at the Single-Molecule Level. Molecular Cell. 2015 May; 58(4):644–659. https://www.sciencedirect.com/science/article/pii/S1097276515001653, doi: 10.1016/j.molcel.2015.02.033.

Lotka AJ. UNDAMPED OSCILLATIONS DERIVED FROM THE LAW OF MASS ACTION. Journal of the American Chemical Society. 1920 Aug; 42(8):1595–1599. https://doi.org/10.1021/ja01453a010, doi: 10.1021/ja01453a010, publisher: American Chemical Society.

Lowrey PL, Shimomura K, Antoch MP, Yamazaki S, Zemenides PD, Ralph MR, Menaker M, Takahashi JS. Positional Syntenic Cloning and Functional Characterization of the Mammalian Circadian Mutation tau. Science. 2000 Apr; 288(5465):483–491. https://doi.org/10.1126/science.288.5465.483, doi: 10.1126/science.288.5465.483, publisher: American Association for the Advancement of Science.

Lévi F, Komarzynski S, Huang Q, Young T, Ang Y, Fuller C, Bolborea M, Brettschneider J, Fursse J, Finkenstädt B, White DP, Innominato P. Tele-Monitoring of Cancer Patients' Rhythms during Daily Life Identifies Actionable Determinants of Circadian and Sleep Disruption. Cancers. 2020; 12(7). doi: 10.3390/cancers12071938.

Lévi F, Okyar A, Dulong S, Innominato PF, Clairambault J. Circadian Timing in Cancer Treatments. Annual Review of Pharmacology and Toxicology. 2010 Feb; 50(1):377–421. https://doi.org/10.1146/annurev.pharmtox.48.113006.094626, doi: 10.1146/annurev.pharmtox.48.113006.094626, publisher: Annual Reviews.

Maarleveld TR, Olivier BG, Bruggeman FJ. StochPy: A Comprehensive, User-Friendly Tool for Simulating Stochastic Biological Processes. PLOS ONE. 2013 Nov; 8(11):1–10. https://doi.org/10.1371/journal.pone.0079345, doi: 10.1371/journal.pone.0079345, publisher: Public Library of Science.

Makoto A, Yoshiki T, Takao Y, Eisuke N. Control of Intracellular Dynamics of Mammalian Period Proteins by Casein Kinase I (CKI) and CKI in Cultured Cells. Molecular and Cellular Biology. 2002 Mar; 22(6):1693–1703. https://doi.org/10.1128/MCB.22.6.1693-1703.2002, doi: 10.1128/MCB.22.6.1693-1703.2002, publisher: American Society for Microbiology.

Malik-Sheriff RS, Glont M, Nguyen TVN, Tiwari K, Roberts MG, Xavier A, Vu MT, Men J, Maire M, Kananathan S, Fairbanks EL, Meyer JP, Arankalle C, Varusai TM, Knight-Schrijver V, Li L, Dueñas-Roca C, Dass G, Keating SM, Park YM, et al. BioModels-15 years of sharing computational models in life science. Nucleic Acids Res. 2020 Jan; 48(D1):D407–D415. Place: England.

Matsuo T, Yamaguchi S, Mitsui S, Emi A, Shimoda F, Okamura H. Control Mechanism of the Circadian Clock for

Timing of Cell Division in Vivo. Science. 2003 Oct; 302(5643):255–259. https://doi.org/10.1126/science.1086271, doi: 10.1126/science.1086271, publisher: American Association for the Advancement of Science.

McManus D, Polidarova L, Smyllie NJ, Patton AP, Chesham JE, Maywood ES, Chin JW, Hastings MH. Cryptochrome 1 as a state variable of the circadian clockwork of the suprachiasmatic nucleus: Evidence from translational switching. Proceedings of the National Academy of Sciences. 2022 Aug; 119(34):e2203563119. https://doi.org/10.1073/pnas.2203563119, doi: 10.1073/pnas.2203563119, publisher: Proceedings of the National Academy of Sciences.

Meng QJ, Logunova L, Maywood ES, Gallego M, Lebiecki J, Brown TM, Sládek M, Semikhodskii AS, Glossop NRJ, Piggins HD, Chesham JE, Bechtold DA, Yoo SH, Takahashi JS, Virshup DM, Boot-Handford RP, Hastings MH, Loudon ASI. Setting Clock Speed in Mammals: The CK1 tau Mutation in Mice Accelerates Circadian Pacemakers by Selectively Destabilizing PERIOD Proteins. Neuron. 2008 Apr; 58(1):78–88. https://doi.org/10.1016/j.neuron.2008.01.019, doi: 10.1016/j.neuron.2008.01.019, publisher: Elsevier.

Merrow M, Roenneberg T, Macino G, Franchi L. A fungus among us: the Neurospora crassa circadian system. Seminars in Cell & Developmental Biology. 2001 Aug; 12(4):279–285. https://www.sciencedirect.com/science/article/pii/S1084952101902551, doi: 10.1006/scdb.2001.0255.

Meyer-Bernstein EL, Jetton AE, Matsumoto Si, Markuns JF, Lehman MN, Bittman EL. Effects of Suprachiasmatic Transplants on Circadian Rhythms of Neuroendocrine Function in Golden Hamsters**This work was supported by NIH Grants MH-44132, KO2-MH-00914, and F32-HD-07673. A preliminary report of this research was presented at the 23rd Annual Meeting of the Society for Neuroscience (Neurosci Abstr 19:236.17, 1993). Endocrinology. 1999 Jan; 140(1):207–218. https://doi.org/10.1210/endo.140.1.6428, doi: 10.1210/endo.140.1.6428.

Michelman-Ribeiro A, Mazza D, Rosales T, Stasevich TJ, Boukari H, Rishi V, Vinson C, Knutson JR, McNally JG. Direct Measurement of Association and Dissociation Rates of DNA Binding in Live Cells by Fluorescence Correlation Spectroscopy. Biophysical Journal. 2009 Jul; 97(1):337–346. https://doi.org/10.1016/j.bpj.2009.04.027, doi: 10.1016/j.bpj.2009.04.027, publisher: Elsevier.

Miller SJ. The method of least squares. Mathematics Department Brown University. 2006; 8:1–7.

Mirjalili S. Genetic Algorithm. In: Mirjalili S, editor. *Evolutionary Algorithms and Neural Networks: Theory and Applications* Cham: Springer International Publishing; 2019.p. 43–55. https://doi.org/10.1007/978-3-319-93025-1_4, doi: 10.1007/978-3-319-93025-1_4.

Mirsky HP, Liu AC, Welsh DK, Kay SA, Doyle FJ. A model of the cell-autonomous mammalian circadian clock. Proceedings of the National Academy of Sciences. 2009 Jul; 106(27):11107–11112. https://doi.org/10.1073/pnas.0904837106, doi: 10.1073/pnas.0904837106, publisher: Proceedings of the National Academy of Sciences.

Mitsui S, Yamaguchi S, Matsuo T, Ishida Y, Okamura H. Antagonistic role of E4BP4 and PAR proteins in the circadian oscillatory mechanism. Genes & Development. 2001 Apr; 15(8):995–1006. http://genesdev.cshlp.org/content/15/8/995.abstract, doi: 10.1101/gad.873501.

Moore RY, Eichler VB. Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. Brain Research. 1972 Jul; 42(1):201–206. https://www.sciencedirect.com/science/article/pii/0006899372900546, doi: 10.1016/0006-8993(72)90054-6.

Motulsky HJ, Ransnas LA. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. The FASEB Journal. 1987 Nov; 1(5):365–374. https://doi.org/10.1096/fasebj.1.5.3315805, doi: 10.1096/fasebj.1.5.3315805, publisher: John Wiley & Sons, Ltd.

Myers MP, Wager-Smith K, Wesley CS, Young MW, Sehgal A. Positional Cloning and Sequence Analysis of the Drosophila Clock Gene, timeless. Science. 1995 Nov; 270(5237):805–808. https://doi.org/10.1126/science.270.5237.805, doi: 10.1126/science.270.5237.805, publisher: American Association for the Advancement of Science.

Müller JD, Chen Y, Gratton E. Resolving Heterogeneity on the Single Molecular Level with the Photon-Counting Histogram. Biophysical Journal. 2000 Jan; 78(1):474–486. https://doi.org/10.1016/S0006-3495(00)76610-0, doi: 10.1016/S0006-3495(00)76610-0, publisher: Elsevier.

Müller P, Schwille P, Weidemann T. PyCorrFit—generic data evaluation for fluorescence correlation spectroscopy. Bioinformatics. 2014 Sep; 30(17):2532–2533. https://doi.org/10.1093/bioinformatics/btu328, doi: 10.1093/bioinformatics/btu328.

Nakajima M, Imai K, Ito H, Nishiwaki T, Murayama Y, Iwasaki H, Oyama T, Kondo T. Reconstitution of Circadian Oscillation of Cyanobacterial KaiC Phosphorylation in Vitro. Science. 2005 Apr; 308(5720):414–415. https://doi.org/10.1126/science.1108451, doi: 10.1126/science.1108451, publisher: American Association for the Advancement of Science.

Nangle SN, Rosensweig C, Koike N, Tei H, Takahashi JS, Green CB, Zheng N. Molecular assembly of the period-cryptochrome circadian transcriptional repressor complex. eLife. 2014 Aug; 3:e03674. https://doi.org/10.7554/eLife.03674, doi: 10.7554/eLife.03674, publisher: eLife Sciences Publications, Ltd.

Narasimamurthy R, Virshup DM. The phosphorylation switch that regulates ticking of the circadian clock. Molecular

Cell. 2021 Mar; 81(6):1133–1146. https://doi.org/10.1016/j.molcel.2021.01.006, doi: 10.1016/j.molcel.2021.01.006, publisher: Elsevier.

Narumi R, Shimizu Y, Ukai-Tadenuma M, Ode KL, Kanda GN, Shinohara Y, Sato A, Matsumoto K, Ueda HR. Mass spectrometry-based absolute quantification reveals rhythmic variation of mouse circadian clock proteins. Proceedings of the National Academy of Sciences. 2016 Jun; 113(24):E3461–E3467. https://doi.org/10.1073/pnas.1603799113, doi: 10.1073/pnas.1603799113, publisher: Proceedings of the National Academy of Sciences.

Nelson DE, Ihekwaba AEC, Elliott M, Johnson JR, Gibney CA, Foreman BE, Nelson G, See V, Horton CA, Spiller DG, Edwards SW, McDowell HP, Unitt JF, Sullivan E, Grimley R, Benson N, Broomhead D, Kell DB, White MRH. Oscillations in NF- B Signaling Control the Dynamics of Gene Expression. Science. 2004 Oct; 306(5696):704–708. https://doi.org/10.1126/science.1099962, doi: 10.1126/science.1099962, publisher: American Association for the Advancement of Science.

Nobel Prize Outreach A, Advanced information. NobelPrize.org.; 2023. https://www.nobelprize.org/prizes/medicine/2017/advanced-information/.

Novák B, Tyson JJ. Design principles of biochemical oscillators. Nature Reviews Molecular Cell Biology. 2008 Dec; 9(12):981–991. https://doi.org/10.1038/nrm2530, doi: 10.1038/nrm2530.

Oishi Y, Hayashi S, Isagawa T, Oshima M, Iwama A, Shimba S, Okamura H, Manabe I. Bmal1 regulates inflammatory responses in macrophages by modulating enhancer RNA transcription. Scientific Reports. 2017 Aug; 7(1):7086. https://doi.org/10.1038/s41598-017-07100-3, doi: 10.1038/s41598-017-07100-3.

Pan X, Bradfield CA, Hussain MM. Global and hepatocyte-specific ablation of Bmal1 induces hyperlipidaemia and enhances atherosclerosis. Nature Communications. 2016 Oct; 7(1):13011. https://doi.org/10.1038/ncomms13011, doi: 10.1038/ncomms13011.

Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, Schultz PG, Kay SA, Takahashi JS, Hogenesch JB. Coordinated Transcription of Key Pathways in the Mouse by the Circadian Clock. Cell. 2002 May; 109(3):307–320. https://www.sciencedirect.com/science/article/pii/S0092867402007225, doi: 10.1016/S0092-8674(02)00722-5.

Panda S, Sato TK, Castrucci AM, Rollag MD, DeGrip WJ, Hogenesch JB, Provencio I, Kay SA. Melanopsin (Opn4) Requirement for Normal Light-Induced Circadian Phase Shifting. Science. 2002 Dec; 298(5601):2213–2216. https://doi.org/10.1126/science.1076848, doi: 10.1126/science.1076848, publisher: American Association for the Advancement of Science.

Peek CB, Levine DC, Cedernaes J, Taguchi A, Kobayashi Y, Tsai SJ, Bonar NA, McNulty MR, Ramsey KM, Bass J. Circadian Clock Interaction with HIF1 Mediates Oxygenic Metabolism and Anaerobic Glycolysis in Skeletal Muscle. Cell Metabolism. 2017 Jan; 25(1):86–92. https://doi.org/10.1016/j.cmet.2016.09.010, doi: 10.1016/j.cmet.2016.09.010, publisher: Elsevier.

Pfanzagl J. Parametric Statistical Theory. De Gruyter; 1994. https://doi.org/10.1515/9783110889765, doi: 10.1515/9783110889765.

Phillips NE, Hugues A, Yeung J, Durandau E, Nicolas D, Naef F. The circadian oscillator analysed at the single-transcript level. Molecular Systems Biology. 2021 Mar; 17(3):e10135. https://doi.org/10.15252/msb.202010135, doi: 10.15252/msb.202010135, publisher: John Wiley & Sons, Ltd.

Phillips NE, Manning CS, Pettini T, Biga V, Marinopoulou E, Stanley P, Boyd J, Bagnall J, Paszek P, Spiller DG, White MR, Goodfellow M, Galla T, Rattray M, Papalopulu N. Stochasticity in the miR-9/Hes1 oscillatory network can account for clonal heterogeneity in the timing of differentiation. eLife. 2016 Oct; 5:e16118. https://doi.org/10.7554/eLife.16118, doi: 10.7554/eLife.16118, publisher: eLife Sciences Publications, Ltd.

Pittendrigh CS. Temporal Organization: Reflections of a Darwinian Clock-Watcher. Annual Review of Physiology. 1993 Oct; 55(1):17–54. https://doi.org/10.1146/annurev.ph.55.030193.000313, doi: 10.1146/annurev.ph.55.030193.000313, publisher: Annual Reviews.

Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the Aequorea victoria green-fluorescent protein. Gene. 1992 Feb; 111(2):229–233. https://www.sciencedirect.com/science/article/pii/037811199290691H, doi: 10.1016/0378-1119(92)90691-H.

Preitner N, Damiola F, Lopez-Molina L, Zakany J, Duboule D, Albrecht U, Schibler U. The Orphan Nuclear Receptor REV-ERB Controls Circadian Transcription within the Positive Limb of the Mammalian Circadian Oscillator. Cell. 2002 Jul; 110(2):251–260. https://doi.org/10.1016/S0092-8674(02)00825-5, doi: 10.1016/S0092-8674(02)00825-5, publisher: Elsevier.

Qian H. On the statistics of fluorescence correlation spectroscopy. Biophysical Chemistry. 1990; 38(1):49–57. https://www.sciencedirect.com/science/article/pii/030146229080039A, doi: https://doi.org/10.1016/0301-4622(90)80039-A.

Ralph MR, Foster RG, Davis FC, Menaker M. Transplanted Suprachiasmatic Nucleus Determines Circadian Period. Science. 1990 Feb; 247(4945):975–978. https://doi.org/10.1126/science.2305266, doi: 10.1126/science.2305266, publisher: American Association for the Advancement of Science.

Ralph MR, Menaker M. A Mutation of the Circadian System in Golden Hamsters. Science. 1988 Sep; 241(4870):1225–1227. https://doi.org/10.1126/science.3413487, doi: 10.1126/science.3413487, publisher: American Association for the Advancement of Science.

Reischl S, Vanselow K, Westermark PO, Thierfelder N, Maier B, Herzel H, Kramer A. -TrCP1-Mediated Degradation of PERIOD2 Is Essential for Circadian Dynamics. Journal of Biological Rhythms. 2007 Oct; 22(5):375–386. https://doi.org/10.1177/0748730407303926, doi: 10.1177/0748730407303926, publisher: SAGE Publications Inc.

Relógio A, Westermark PO, Wallach T, Schellenberg K, Kramer A, Herzel H. Tuning the Mammalian Circadian Clock: Robust Synergy of Two Loops. PLOS Computational Biology. 2011 Dec; 7(12):e1002309. https://doi.org/10.1371/journal.pcbi.1002309, doi: 10.1371/journal.pcbi.1002309, publisher: Public Library of Science.

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 Biology. 2011 Feb; 9(2):1–18. https://doi.org/10.1371/journal.pbio.1000595, doi: 10.1371/journal.pbio.1000595, publisher: Public Library of Science.

Ruben MD, Smith DF, FitzGerald GA, Hogenesch JB. Dosing time matters. Science. 2019 Aug; 365(6453):547–549. https://doi.org/10.1126/science.aax7621, doi: 10.1126/science.aax7621, publisher: American Association for the Advancement of Science.

Sadaie W, Harada Y, Matsuda M, Aoki K. Quantitative In Vivo Fluorescence Cross-Correlation Analyses Highlight the Importance of Competitive Effects in the Regulation of Protein-Protein Interactions. Molecular and Cellular Biology. 2014; 34(17):3272–3290. https://mcb.asm.org/content/34/17/3272, doi: 10.1128/MCB.00087-14, publisher: American Society for Microbiology Journals _eprint: https://mcb.asm.org/content/34/17/3272.full.pdf.

Saffarian S, Elson EL. Statistical Analysis of Fluorescence Correlation Spectroscopy: The Standard Deviation and Bias. Biophysical Journal. 2003; 84(3):2030–2042. https://www.sciencedirect.com/science/article/pii/S0006349503750115, doi: https://doi.org/10.1016/S0006-3495(03)75011-5.

Sato TK, Panda S, Miraglia LJ, Reyes TM, Rudic RD, McNamara P, Naik KA, FitzGerald GA, Kay SA, Hogenesch JB. A Functional Genomics Strategy Reveals Rora as a Component of the Mammalian Circadian Clock. Neuron. 2004 Aug; 43(4):527–537. https://doi.org/10.1016/j.neuron.2004.07.018, doi: 10.1016/j.neuron.2004.07.018, publisher: Elsevier.

Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, et al. Database resources of the national center for biotechnology information. Nucleic Acids Research. 2021 Dec; 50(D1):D20–D26. https://doi.org/10.1093/nar/gkab1112, doi: 10.1093/nar/gkab1112, _eprint: https://academic.oup.com/nar/article-pdf/50/D1/D20/42058080/gkab1112.pdf.

Scheiermann C, Gibbs J, Ince L, Loudon A. Clocking in to immunity. Nature Reviews Immunology. 2018 Jul; 18(7):423–437. https://doi.org/10.1038/s41577-018-0008-4, doi: 10.1038/s41577-018-0008-4.

Schmalen I, Reischl S, Wallach T, Klemz R, Grudziecki A, Prabu JR, Benda C, Kramer A, Wolf E. Interaction of circadian clock proteins CRY1 and PER2 is modulated by zinc binding and disulfide bond formation. Cell. 2014; 157(5):1203–15. https://www.ncbi.nlm.nih.gov/pubmed/24855952, doi: 10.1016/j.cell.2014.03.057, type: Journal Article.

Schmidt HG, Sewitz S, Andrews SS, Lipkow K. An Integrated Model of Transcription Factor Diffusion Shows the Importance of Intersegmental Transfer and Quaternary Protein Structure for Target Site Finding. PLOS ONE. 2014 Oct; 9(10):e108575. https://doi.org/10.1371/journal.pone.0108575, doi: 10.1371/journal.pone.0108575, publisher: Public Library of Science.

Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics. 1978; 6(2):461–464. http://www.jstor.org/stable/2958889, publisher: Institute of Mathematical Statistics.

Schwille P, Bieschke J, Oehlenschläger F. Kinetic investigations by fluorescence correlation spectroscopy: The analytical and diagnostic potential of diffusion studies. Biophysical Chemistry. 1997; 66(2):211–228. https://www.sciencedirect.com/science/article/pii/S0301462297000616, doi: https://doi.org/10.1016/S0301-4622(97)00061-6.

Schätzel K, Drewel M, Stimac S. Photon Correlation Measurements at Large Lag Times: Improving Statistical Accuracy. Journal of Modern Optics. 1988; 35(4):711–718. https://doi.org/10.1080/09500348814550731, doi: 10.1080/09500348814550731, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09500348814550731.

Sehgal A, Rothenfluh-Hilfiker A, Hunter-Ensor M, Chen Y, Myers MP, Young MW. Rhythmic Expression of timeless: A Basis for Promoting Circadian Cycles in period Gene Autoregulation. Science. 1995 Nov; 270(5237):808–810. https://doi.org/10.1126/science.270.5237.808, doi: 10.1126/science.270.5237.808, publisher: American Association for the Advancement of Science.

Shah M, Funnell APW, Quinlan KGR, Crossley M. Hit and Run Transcriptional Repressors Are Difficult to Catch in the Act. BioEssays. 2019 Aug; 41(8):1900041. https://doi.org/10.1002/bies.201900041, doi: 10.1002/bies.201900041, publisher: John Wiley & Sons, Ltd.

Shamir M, Bar-On Y, Phillips R, Milo R. SnapShot: Timescales in Cell Biology. Cell. 2016 Mar; 164(6):1302–1302.e1. https://www.sciencedirect.com/science/article/pii/S0092867416302082, doi: 10.1016/j.cell.2016.02.058.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research. 2003; 13(11):2498–2504.

http://genome.cshlp.org/content/13/11/2498.abstract, doi: 10.1101/gr.1239303, _eprint: http://genome.cshlp.org/content/13/11/2498.full.pdf+html.

Shao J. AN ASYMPTOTIC THEORY FOR LINEAR MODEL SELECTION. Statistica Sinica. 1997; 7(2):221–242. http://www.jstor.org/stable/24306073, publisher: Institute of Statistical Science, Academia Sinica.

Shearman LP, Xiaowei J, Lee C, Reppert SM, Weaver DR. Targeted Disruption of the mPer3 Gene: Subtle Effects on Circadian Clock Function. Molecular and Cellular Biology. 2000 Sep; 20(17):6269–6275. https://doi.org/10.1128/MCB.20.17.6269-6275.2000, doi: 10.1128/MCB.20.17.6269-6275.2000, publisher: American Society for Microbiology.

Shirogane T, Jin J, Ang XL, Harper JW. SCF -TRCP Controls Clock-dependent Transcription via Casein Kinase 1-dependent Degradation of the Mammalian Period-1 (Per1) Proteinm *. Journal of Biological Chemistry. 2005 Jul; 280(29):26863–26872. https://doi.org/10.1074/jbc.M502862200, doi: 10.1074/jbc.M502862200, publisher: Elsevier.

Siwicki KK, Eastman C, Petersen G, Rosbash M, Hall JC. Antibodies to the period gene product of drosophila reveal diverse tissue distribution and rhythmic changes in the visual system. Neuron. 1988 Apr; 1(2):141–150. https://doi.org/10.1016/0896-6273(88)90198-5, doi: 10.1016/0896-6273(88)90198-5, publisher: Elsevier.

Skellam JG. The Frequency Distribution of the Difference Between Two Poisson Variates Belonging to Different Populations. Journal of the Royal Statistical Society. 1946; 109(3):296–296. http://www.jstor.org/stable/2981372, doi: 10.2307/2981372, publisher: [Wiley, Royal Statistical Society].

von Smoluchowski M. Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. Annalen der Physik. 1906 Jan; 326(14):756–780. https://doi.org/10.1002/andp.19063261405, doi: 10.1002/andp.19063261405, publisher: John Wiley & Sons, Ltd.

Smyllie NJ, Bagnall J, Koch AA, Niranjan D, Polidarova L, Chesham JE, Chin JW, Partch CL, Loudon ASI, Hastings MH. Cryptochrome proteins regulate the circadian intracellular behavior and localization of PER2 in mouse suprachiasmatic nucleus neurons. Proceedings of the National Academy of Sciences. 2022 Jan; 119(4):e2113845119. https://doi.org/10.1073/pnas.2113845119, doi: 10.1073/pnas.2113845119, publisher: Proceedings of the National Academy of Sciences.

Smyllie N, Pilorz V, Boyd J, Meng QJ, Saer B, Chesham J, Maywood E, Krogager T, Spiller D, Boot-Handford R, White MH, Hastings M, Loudon AI. Visualizing and Quantifying Intracellular Behavior and Abundance of the Core Circadian Clock Protein PERIOD2. Current Biology. 2016 Jul; 26(14):1880–1886. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4963210/, doi: 10.1016/j.cub.2016.05.018, number: 14.

Spiller DG, Wood CD, Rand DA, White MRH. Measurement of single-cell dynamics. Nature. 2010 Jun; 465(7299):736–745. https://doi.org/10.1038/nature09232, doi: 10.1038/nature09232.

Sprague BL, McNally JG. FRAP analysis of binding: proper and fitting. Trends in Cell Biology. 2005 Feb; 15(2):84–91. https://www.sciencedirect.com/science/article/pii/S0962892404003332, doi: 10.1016/j.tcb.2004.12.001.

Sprague BL, Pego RL, Stavreva DA, McNally JG. Analysis of Binding Reactions by Fluorescence Recovery after Photobleaching. Biophysical Journal. 2004 Jun; 86(6):3473–3495. https://doi.org/10.1529/biophysj.103.026765, doi: 10.1529/biophysj.103.026765, publisher: Elsevier.

Stephan FK, Zucker I. Circadian Rhythms in Drinking Behavior and Locomotor Activity of Rats Are Eliminated by Hypothalamic Lesions. Proceedings of the National Academy of Sciences. 1972 Jun; 69(6):1583–1586. https://doi.org/10.1073/pnas.69.6.1583, doi: 10.1073/pnas.69.6.1583, publisher: Proceedings of the National Academy of Sciences.

Stone M. An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. Journal of the Royal Statistical Society Series B (Methodological). 1977; 39(1):44–47. http://www.jstor.org/stable/2984877, publisher: [Royal Statistical Society, Wiley].

Storch KF, Lipan O, Leykin I, Viswanathan N, Davis FC, Wong WH, Weitz CJ. Extensive and divergent circadian gene expression in liver and heart. Nature. 2002 May; 417(6884):78–83. https://doi.org/10.1038/nature744, doi: 10.1038/nature744.

Stratmann M, Suter DM, Molina N, Naef F, Schibler U. Circadian Dbp transcription relies on highly dynamic BMAL1-CLOCK interaction with E boxes and requires the proteasome. Mol Cell. 2012; 48(2):277–87. https://www.ncbi.nlm.nih.gov/pubmed/22981862, doi: 10.1016/j.molcel.2012.08.012, type: Journal Article.

Strogatz SH. Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. CRC Press; 2018. https://books.google.co.uk/books?id=A0paDwAAQBAJ.

Stuart AM. Inverse problems: A Bayesian perspective. Acta Numerica. 2010; 19:451–559. https://www.cambridge.org/core/article/inverse-problems-a-bayesian-perspective/587A3A0D480A1A7C2B1B284BCEDF7E23, doi: 10.1017/S0962492910000061, edition: 2010/05/10 Publisher: Cambridge University Press.

Takahashi JS. Transcriptional architecture of the mammalian circadian clock. Nature Reviews Genetics. 2017 Mar; 18(3):164–179. https://doi.org/10.1038/nrg.2016.150, doi: 10.1038/nrg.2016.150.

Tavakoli M, Jazani S, Sgouralis I, Shafraz OM, Sivasankar S, Donaphon B, Levitus M, Pressé S. Pitching Single-Focus Confocal Data Analysis One Photon at a Time with Bayesian Nonparametrics. Phys Rev X. 2020 Jan; 10(1):011021. https://link.aps.org/doi/10.1103/PhysRevX.10.011021, doi: 10.1103/PhysRevX.10.011021, publisher: American Physical Society.

The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research. 2022 Nov; 51(D1):D523–D531. https://doi.org/10.1093/nar/gkac1052, doi: 10.1093/nar/gkac1052, _eprint: https://academic.oup.com/nar/article-pdf/51/D1/D523/48441158/gkac1052.pdf.

Tian Y, Martinez MM, Pappas D. Fluorescence Correlation Spectroscopy: A Review of Biochemical and Microfluidic Applications. Applied Spectroscopy. 2011 Apr; 65(4):115A–115A. https://journals.sagepub.com/doi/abs/10.1366/10-06224, doi: 10.1366/10-06224, publisher: SAGE Publications Ltd STM.

Tiwari J, Fraser A. Genetic regulation by feedback repression. Journal of Theoretical Biology. 1973 Jun; 39(3):679–681. https://www.sciencedirect.com/science/article/pii/0022519373900829, doi: 10.1016/0022-5193(73)90082-9.

Toh KL, Jones CR, He Y, Eide EJ, Hinz WA, Virshup DM, Ptáček LJ, Fu YH. An hPer2 Phosphorylation Site Mutation in Familial Advanced Sleep Phase Syndrome. Science. 2001 Feb; 291(5506):1040–1043. https://doi.org/10.1126/science.1057499, doi: 10.1126/science.1057499, publisher: American Association for the Advancement of Science.

Trott AJ, Menet JS. Regulation of circadian clock transcriptional output by CLOCK:BMAL1. PLOS Genetics. 2018 Jan; 14(1):e1007156. https://doi.org/10.1371/journal.pgen.1007156, doi: 10.1371/journal.pgen.1007156, publisher: Public Library of Science.

Tsai TYC, Choi YS, Ma W, Pomerening JR, Tang C, Ferrell JE. Robust, Tunable Biological Oscillations from Interlinked Positive and Negative Feedback Loops. Science. 2008 Jul; 321(5885):126–129. https://doi.org/10.1126/science.1156951, doi: 10.1126/science.1156951, publisher: American Association for the Advancement of Science.

Ueda HR, Hayashi S, Chen W, Sano M, Machida M, Shigeyoshi Y, Iino M, Hashimoto S. System-level identification of transcriptional circuits underlying mammalian circadian clocks. Nature Genetics. 2005 Feb; 37(2):187–192. https://doi.org/10.1038/ng1504, doi: 10.1038/ng1504.

Unosson M, Brancaccio M, Hastings M, Johansen AM, Finkenstädt B. A spatio-temporal model to reveal oscillator phenotypes in molecular clocks: Parameter estimation elucidates circadian gene transcription dynamics in single-cells. PLOS Computational Biology. 2021 Dec; 17(12):e1009698. https://doi.org/10.1371/journal.pcbi.1009698, doi: 10.1371/journal.pcbi.1009698, publisher: Public Library of Science.

Vanselow K, Vanselow JT, Westermark PO, Reischl S, Maier B, Korte T, Herrmann A, Herzel H, Schlosser A, Kramer A. Differential effects of PER2 phosphorylation: molecular basis for the human familial advanced sleep phase syndrome (FASPS). Genes & Development. 2006 Oct; 20(19):2660–2672. http://genesdev.cshlp.org/content/20/19/2660.abstract.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Žídek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research. 2021 Nov; 50(D1):D439–D444. https://doi.org/10.1093/nar/gkab1061, doi: 10.1093/nar/gkab1061, _eprint: https://academic.oup.com/nar/article-pdf/50/D1/D439/43502749/gkab1061.pdf.

van der Veen DR, Shao J, Xi Y, Li L, Duffield GE. Cardiac Atrial Circadian Rhythms in PERIOD2::LUCIFERASE and per1:luc Mice: Amplitude and Phase Responses to Glucocorticoid Signaling and Medium Treatment. PLOS ONE. 2012 Oct; 7(10):e47692. https://doi.org/10.1371/journal.pone.0047692, doi: 10.1371/journal.pone.0047692, publisher: Public Library of Science.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020; 17:261–272. doi: 10.1038/s41592-019-0686-2.

Vitaterna MH, King DP, Chang AM, Kornhauser JM, Lowrey PL, McDonald JD, Dove WF, Pinto LH, Turek FW, Takahashi JS. Mutagenesis and Mapping of a Mouse Gene, Clock, Essential for Circadian Behavior. Science. 1994 Apr; 264(5159):719–725. https://doi.org/10.1126/science.8171325, doi: 10.1126/science.8171325, publisher: American Association for the Advancement of Science.

Voit EO, Martens HA, Omholt SW. 150 Years of the Mass Action Law. PLOS Computational Biology. 2015 Jan; 11(1):e1004012. https://doi.org/10.1371/journal.pcbi.1004012, doi: 10.1371/journal.pcbi.1004012, publisher: Public Library of Science.

Wachsmuth M, Conrad C, Bulkescher J, Koch B, Mahen R, Isokane M, Pepperkok R, Ellenberg J. High-throughput fluorescence correlation spectroscopy enables analysis of proteome dynamics in living cells. Nature Biotechnology. 2015 Apr; 33(4):384–389. https://doi.org/10.1038/nbt.3146, doi: 10.1038/nbt.3146.

Wang GZ, Hickey S, Shi L, Huang HC, Nakashe P, Koike N, Tu B, Takahashi J, Konopka G. Cycling Transcriptional Networks Optimize Energy Utilization on a Genome Scale. Cell Reports. 2015 Dec; 13(9):1868–1880. https://doi.org/10.1016/j.celrep.2015.10.043, doi: 10.1016/j.celrep.2015.10.043, publisher: Elsevier.

Webb AB, Angelo N, Huettner JE, Herzog ED. Intrinsic, nondeterministic circadian rhythm generation in identified mammalian neurons. Proceedings of the National Academy of Sciences. 2009 Sep; 106(38):16493–16498. https://doi.org/10.1073/pnas.0902768106, doi: 10.1073/pnas.0902768106, publisher: Proceedings of the National Academy of Sciences.

Webb AB, Taylor SR, Thoroughman KA, Doyle FJ III, Herzog ED. Weakly Circadian Cells Improve Resynchrony. PLOS Computational Biology. 2012 Nov; 8(11):e1002787. https://doi.org/10.1371/journal.pcbi.1002787, doi: 10.1371/journal.pcbi.1002787, publisher: Public Library of Science.

Weiss NA. A course in probability. Upper Saddle River, NJ: Pearson; 2005.

Welsh DK, Yoo SH, Liu AC, Takahashi JS, Kay SA. Bioluminescence Imaging of Individual Fibroblasts Reveals Persistent, Independently Phased Circadian Rhythms of Clock Gene Expression. Current Biology. 2004 Dec; 14(24):2289–2295. https://doi.org/10.1016/j.cub.2004.11.057, doi: 10.1016/j.cub.2004.11.057, publisher: Elsevier.

Wichert S, Fokianos K, Strimmer K. Identifying periodically expressed transcripts in microarray time series data. Bioinformatics. 2004 Jan; 20(1):5–20. https://doi.org/10.1093/bioinformatics/btg364, doi: 10.1093/bioinformatics/btg364.

Widengren J, Mets U, Rigler R. Fluorescence correlation spectroscopy of triplet states in solution: a theoretical and experimental study. The Journal of Physical Chemistry. 1995 Sep; 99(36):13368–13379. https://doi.org/10.1021/j100036a009, doi: 10.1021/j100036a009, publisher: American Chemical Society.

Woller A, Gonze D, Erneux T. The Goodwin model revisited: Hopf bifurcation, limit-cycle, and periodic entrainment. Physical Biology. 2014 Jul; 11(4):045002. http://dx.doi.org/10.1088/1478-3975/11/4/045002, doi: 10.1088/1478-3975/11/4/045002, publisher: IOP Publishing.

Wong DC, O'Neill JS. Non-transcriptional processes in circadian rhythm generation. Circadian Rhythms. 2018 Oct; 5:117–132. https://www.sciencedirect.com/science/article/pii/S2468867318301342, doi: 10.1016/j.cophys.2018.10.003.

Wood SN. Institute of mathematical statistics textbooks: Core statistics series number 6. Cambridge, England: Cambridge University Press; 2015.

Wu Y, Tang D, Liu N, Xiong W, Huang H, Li Y, Ma Z, Zhao H, Chen P, Qi X, Zhang EE. Reciprocal Regulation between the Circadian Clock and Hypoxia Signaling at the Genome Level in Mammals. Cell Metabolism. 2017 Jan; 25(1):73–85. https://doi.org/10.1016/j.cmet.2016.09.009, doi: 10.1016/j.cmet.2016.09.009, publisher: Elsevier.

Xie Y, Tang Q, Chen G, Xie M, Yu S, Zhao J, Chen L. New Insights Into the Circadian Rhythm and Its Related Diseases. Frontiers in Physiology. 2019; 10. https://www.frontiersin.org/articles/10.3389/fphys.2019.00682.

Xu H, Gustafson CL, Sammons PJ, Khan SK, Parsley NC, Ramanathan C, Lee HW, Liu AC, Partch CL. Cryptochrome 1 regulates the circadian clock through dynamic interactions with the BMAL1 C terminus. Nat Struct Mol Biol. 2015; 22(6):476–484. https://www.ncbi.nlm.nih.gov/pubmed/25961797, doi: 10.1038/nsmb.3018, type: Journal Article.

Xu Y, Padiath QS, Shapiro RE, Jones CR, Wu SC, Saigoh N, Saigoh K, Ptáček LJ, Fu YH. Functional consequences of a CKI mutation causing familial advanced sleep phase syndrome. Nature. 2005 Mar; 434(7033):640–644. https://doi.org/10.1038/nature03453, doi: 10.1038/nature03453.

Yang N, Smyllie NJ, Morris H, Gonçalves CF, Dudek M, Pathiranage DRJ, Chesham JE, Adamson A, Spiller DG, Zindy E, Bagnall J, Humphreys N, Hoyland J, Loudon ASI, Hastings MH, Meng QJ. Quantitative live imaging of Venus::BMAL1 in a mouse model reveals complex dynamics of the master circadian clock regulator. PLOS Genetics. 2020 Apr; 16(4):e1008729. https://doi.org/10.1371/journal.pgen.1008729, doi: 10.1371/journal.pgen.1008729, publisher: Public Library of Science.

Yang R, Su Z. Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. Bioinformatics. 2010 Jun; 26(12):i168–i174. https://doi.org/10.1093/bioinformatics/btq189, doi: 10.1093/bioinformatics/btq189.

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):1989–98. https://www.ncbi.nlm.nih.gov/pubmed/25228643, doi: 10.1101/gad.249417.114, type: Journal Article.

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):25891–902. https://www.ncbi.nlm.nih.gov/pubmed/21613214, doi: 10.1074/jbc.M111.254680, type: Journal Article.

Yoo SH, Ko CH, Lowrey PL, Buhr ED, Song Ej, Chang S, Yoo OJ, Yamazaki S, Lee C, Takahashi JS. A noncanonical E-box enhancer drives mouse Period2 circadian oscillations in vivo. Proceedings of the National Academy of Sciences. 2005 Feb; 102(7):2608–2613. https://doi.org/10.1073/pnas.0409763102, doi: 10.1073/pnas.0409763102, publisher: Proceedings of the National Academy of Sciences.

Yoo SH, Shin Y, Lowrey PL, Kazuhiro S, H KC, Buhr Ethan D , Siepka Sandra M , Hong Hee-Kyung, Oh Won Jun, Yoo Ook Joon, Menaker Michael, Takahashi Joseph S . PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proceedings of the National Academy of Sciences. 2004 Apr; 101(15):5339–5346. https://doi.org/10.1073/pnas.0308709101, doi: 10.1073/pnas.0308709101, publisher: Proceedings of the National Academy of Sciences.

Young MW, Kay SA. Time zones: a comparative genetics of circadian clocks. Nature Reviews Genetics. 2001 Sep; 2(9):702–715. https://doi.org/10.1038/35088576, doi: 10.1038/35088576.

Yuan Q, Metterville D, Briscoe AD, Reppert SM. Insect Cryptochromes: Gene Duplication and Loss Define Diverse Ways

to Construct Insect Circadian Clocks. Molecular Biology and Evolution. 2007 Apr; 24(4):948–955. https://doi.org/10.1093/molbev/msm011, doi: 10.1093/molbev/msm011.

Zehring WA, Wheeler DA, Reddy P, Konopka RJ, Kyriacou CP, Rosbash M, Hall JC. P-element transformation with period locus DNA restores rhythmicity to mutant, arrhythmic drosophila melanogaster. Cell. 1984 Dec; 39(2, Part 1):369–376. https://www.sciencedirect.com/science/article/pii/0092867484900151, doi: 10.1016/0092-8674(84)90015-1.

Zeng H, Hardin PE, Rosbash M. Constitutive overexpression of the Drosophila period protein inhibits period mRNA cycling. The EMBO Journal. 1994 Aug; 13(15):3590–3598. https://doi.org/10.1002/j.1460-2075.1994.tb06666.x, doi: 10.1002/j.1460-2075.1994.tb06666.x, publisher: John Wiley & Sons, Ltd.

Zhang Y, Fang B, Emmett MJ, Damle M, Sun Z, Feng D, Armour SM, Remsberg JR, Jager J, Soccio RE, Steger DJ, Lazar MA. Discrete functions of nuclear receptor Rev-erb couple metabolism to the clock. Science. 2015 Jun; 348(6242):1488–1492. https://doi.org/10.1126/science.aab3021, doi: 10.1126/science.aab3021, publisher: American Association for the Advancement of Science.

Zhou M, Kim J, Eng G, Forger D, Virshup D. A Period2 Phosphoswitch Regulates and Temperature Compensates Circadian Period. Molecular Cell. 2015 Oct; 60(1):77–88. https://doi.org/10.1016/j.molcel.2015.08.022, doi: 10.1016/j.molcel.2015.08.022, publisher: Elsevier.