Generalization and Discovery by Assuming Conserved Mechanisms: 
Cross Species Research on Circadian Oscillators

William Bechtel

Abstract
In many domains of biology, explanation takes the form of characterizing the mechanism responsible for a particular phenomenon. Mechanisms are often discovered in specific biological systems and this raises the question of how explanations generalize. Generalization often proceeds through recognizing conservation of mechanisms through evolutionary descent. Identifying conservation, though, does more than allow generalization—it serves as a heuristic for discovery since conserved mechanisms are exhibit important variations. This paper illustrates the heuristic functions of assuming conservation of mechanisms for both generalization and discovery by examining recent research into the circadian clock mechanisms of *Drosophila* and mammals.

1. Introduction

How do scientific explanations generalize? When explanation is viewed as the application of scientific laws, generalization is relatively straightforward: laws are universally quantified, and apply to any condition in which the antecedent is satisfied. But biologists, especially in domains such as cell and molecular biology, seldom invoke laws in their explanations. Instead, they describe the mechanisms that they claim are responsible for the phenomenon being explained. Several philosophical accounts of mechanisms and mechanistic explanations in biology have been advanced recently. Although the terminology varies somewhat across authors, the key elements of a mechanistic explanation are the identification of the relevant parts of the mechanism, the determination of the operations they perform, and an account of how the parts and operations are organized and orchestrated so that, under specific contextual conditions, the mechanism realizes the phenomenon of interest (Bechtel and Richardson 1993; Bechtel 2006; Craver 2007; Darden 2006; Machamer, Darden, and Craver 2000). The challenge for understanding generalization is that the mechanisms actually studied are highly particularized: researchers study mechanisms in model systems and it is anticipated that there will be important differences—involving parts, operations, organization, and orchestration—between the mechanisms found in a studied system and those to which one desires to apply the explanation.

A first part of the answer to the question of how mechanistic explanations generalize is that biologists expect there to be similarities among the mechanisms responsible for the same or similar phenomena in different systems. Things can be similar in many respects, though, so just appealing to similarity without appropriate constraint is uninformative. What one often finds in biological practice are appeals to mechanisms whose parts, operations, and organization are conserved though evolutionary descent. The process of descent, however, also introduces variations. Thus, when researchers seek the assumed homologues of the mechanism in the target species they must investigate whether the parts are still operating in the same or sufficiently similar manner and that the overall organization is sufficiently preserved. This heuristic of seeking homologues becomes particularly productive when variants are identified within a mechanism that is still largely conserved. If the variation alters how an operation, other changes are also required for the mechanism to account for the phenomenon, and researchers are guided by their developing understanding of the mechanism to look for specific these other changes.
In this paper I will illustrate how the assumption of conservation served both generalization and discovery in a particular case: the conservation of circadian oscillators within animals. Circadian rhythms, the approximately 24 hour rhythms that are endogenously maintained, are widespread in living systems and affect such aspects as body temperature, metabolism, endocrine function, locomotion, sexual activity, and mental function. Since circadian rhythms are found not only in organisms with a central nervous system but also in single-cell organisms, researchers generally assumed that the internal clock or oscillator employed intracellular processes even in higher organisms. In *Drosophila*, research focused on lateral neurons whereas in mammals it concentrated on a structure in the hypothalamus above the optic chiasm known as the *suprachiasmatic nucleus* (SCN). The explanatory challenge was to explain how concentrations of proteins in these structures could oscillate with a period of approximately 24 hours. In the next sections I consider how research first addressed this challenge using *Drosophila* as the model system and how the assumption that the mechanism was conserved resulted in a productive interaction with research on mammals.

2. The *Drosophila* Model System for Studying Circadian Oscillators

Before the quest to identify the responsible mechanism, *Drosophila* had provided a model for studying circadian behavior due to the fact that fruit flies emerge (eclose) from the pupa only around dawn and even if development is complete, delay ecelosing until the subsequent dawn. Even if kept in total darkness the timing is tightly controlled (Pittendrigh 1954). Fruit flies had also served as a model system for genetic research, so they were plausible target species for Konopka and Benzer (1971) to employ in attempting to identify genes affecting circadian behavior. Applying mutagens targeted to the X-chromosome of *Drosophila*, they generated mutant flies that exhibited shortened, lengthened, and null oscillations in timing of ecellosion. They traced all three mutations to a common locus and named the affect gene *period* (*per*).

The cloning of *per* in the mid-1980s by Michael Rosbash and his colleagues made possible more detailed analysis of its contribution to circadian rhythms. They found the protein synthesized from *per*, PERIOD (PER), is present, often in the nucleus, in a variety of cell types including lateral neurons in the brain as well as in eyes, antennae, and gut. (Protein names are standardly capitalized while the names of genes are written in italics. In *Drosophila* the whole names of genes are in lowercase, whereas in mammals the first letter is capitalized.) Hardin, Hall, and Rosbash (1990) determined that the concentrations of both *per* mRNA and PER exhibited circadian oscillations (with the peaks and valleys in PER concentrations following those of *per* mRNA by approximately 8 hours). Moreover, these cycles were shortened, lengthened, or abolished in the previously discovered *per* mutants. Hardin et al. proposed a feedback mechanism in which PER regulated the transcription of its own gene, *per*. (Figure 1). When there is limited concentration of PER in the nucleus, the synthesis of PER proceeds normally, but as PER accumulates and is transported back into the nucleus, synthesis decreases. As nuclear PER is broken down, synthesis again increases. With appropriate timing of these processes, a regular oscillation in PER concentrations results.
The proposal, though, raised a number of questions about just how the feedback loop operated. Particularly important were the questions of what delayed PER’s translocation back to the nucleus and how PER could inhibit *per* transcription. Researchers realized PER could not do this directly as it lacked a site where it could bind to DNA. This led researchers to search for other component parts in the clock mechanism.

A second part of the mechanism was discovered by Seghal, Price, Man, and Young (1994). They followed essentially the same approach as Konopka, generating a large number of mutations involving the second and third chromosomes in *Drosophila*. The found a mutant that failed to exhibit rhythmic eclosion or locomotion. In these mutants, moreover, *per* mRNA concentrations ceased to oscillate. They took this to indicate an interaction between *per* and the new gene, which they named *timeless* (*tim*). A second study (Vosshall et al. 1994) indicated that *tim* is required for PER to be transported from the cytoplasm, where it is synthesized, back into the nucleus, where it can regulate transcription of its own gene and that “PER contains sequences that somehow inhibit PER nuclear localization in the absence of *tim*” (p. 1607). It was soon determined that PER and TIM form a dimer (a compound of two similar units) before both are transported back into the nucleus (Gekakis et al. 1995) and that a region found on both PER and TIM, which they named the *cytoplasmic localization domain* (CLD), was responsible for preventing either alone from migrating into the nucleus. This region was presumably masked in the dimer, allowing the dimer to be transported into the nucleus (Saez and Young 1996).

The discovery of *tim* helped explain the translocation of PER into the nucleus, but it did not explain how *per* (or *tim*) transcription was inhibited since TIM also lacked a DNA binding site. Investigators next explored the possibility that PER bound with a *per* activating factor, and when it did so, blocked the activator from promoting *per* transcription. Support for the existence of such a factor was provided by the identification of an E-box (CACG TG) upstream of *per* that is a target for a basic-helix-loop-helix (bHLH) DNA binding site, and by determination that it was required for robust cycling (Hao, Allen, and Hardin 1997). The activating factor, however, was first discovered not in *Drosophila* but in mammals.

3. Linking *Drosophila* and mammalian clocks
In the early 1990s the search for circadian mutants was also proceeding on mammals, especially mice. Vitaterna et al. (1994) identified a mutation that yielded increased period length and, in a homozygotic form, loss of rhythms after two weeks. They labeled the mutant gene Clock (for circadian locomotor output cycles kaput) and localized it to chromosome 5. Three years later the same laboratory (King et al. 1997) cloned Clock and predicted “that this candidate gene encodes a novel member of the bHLH–PAS domain family of transcription factors” (p. 645). Noting the assumption of Drosophila researchers that unknown transcription factors interact with PER, King et al. proposed “The mouse Clock gene could be the mammalian ortholog of such a gene” (p. 649). In addition, Gekakis et al. (1998) predicted that CLOCK must dimerize with a partner, and identified BMAL1, a protein whose function was then unknown, as a protein that exhibited a similar pattern of expression as CLOCK. They also demonstrated that the mammalian CLOCK-BMAL1 dimer would bind with Drosophila per’s E-box and proposed processes by which PER might interfere with the action of CLOCK-BMAL1 to inhibit per expression.

The discovery of Clock in mice led Drosophila researchers to seek a Drosophila homolog, and they soon found a gene that possessed very similar bHLH and PAS domains (Darlington et al. 1998). Moreover, they determined that dCLOCK was a specific activator of the per and tim promoters. They also found evidence for a Drosophila homolog of Bmal1 and concluded “It is tempting to speculate that the Drosophila four-component transcriptional feedback loop described here is sufficient to generate a rudimentary circadian rhythm” (p. 1602).

As this work was proceeding, other researchers followed up on an indication of a mammalian homologue to per and in 1997 identified such a homolog to per in mice and in humans. The mouse homolog is approximately 44% identical to the fly protein, with many of the differences involving neutral amino acid substitutions (Sun et al. 1997; Tei et al. 1997). Soon after it was recognized that in fact there are multiple mammalian homologues of per, designated mPer1, mPer2 (Albrecht et al. 1997), and mPer3 (Zylka et al. 1998). They reside on different chromosomes and differ in such ways as their responses to light pulses. One of the first noted consequences of the differentiation of three Per genes in mammals is increased robustness: deletion of just one Per gene does not eliminate circadian cycling, whereas deletion of both Per1 and Per2 does.

4. Just How Conserved is the Mammalian Clock?

At this point there seemed to be a high degree of conservation between the Drosophila and mammalian clock mechanism. Mammalian homologs had been found for per, clock, and bmal1; the only major change was that per had differentiated into three mammalian genes. This supported a straight-forward generalization of the explanation for Drosophila to mammals. But the story soon became more complex as researchers sought a mammalian homolog for the fourth component of the Drosophila clock, tim. They found such a gene whose protein is expressed in the SCN and could dimerize with mPER1 or mPER2 as well as with dPER. In Drosophila cells the dimers were transported back into the nucleus and in mouse cells the dimer with mPER1 inhibited activation of the mPer1 promoter (Sangoram et al. 1998). So far the parallels held up. But there were significant differences: neither the concentrations of mTim mRNA nor its protein, mTIM, oscillate under constant light or darkness, and mTIM levels increase in response to light whereas dTIM levels decrease.
These seemingly minor differences took on greater significance in the context of a major feature of circadian rhythms, their entrainment (resetting) by exposure to light, especially around subjective dawn or subjective dusk. This process is essential for organisms to adjust to seasonal differences or for us to adjust after travel to different time zones. The discovery of photoreceptors in *Drosophila* that are conserved from blue-light photoreceptors (known as *cryptochromes*) in plants provided a critical clue to the entrainment mechanism. Searching for animal homologues, Todo et al. (1996) and Hsu et al. (1996) found one cryptochrome in *Drosophila* and two each in mice and humans. Unlike in plants, concentrations of cryptochromes in *Drosophila* and in mammals were found to undergo circadian oscillations, indicating they might have significant clock functions. In *Drosophila*, Emery et al. (1998) determined that this oscillation was due to light exposure by showing that when flies were maintained in continuous darkness, CRY levels did not oscillate but rather continued to increase. They inferred “CRY is a major photoreceptor for *Drosophila* locomotor activity rhythms” (p. 674). Since TIM concentrations are responsive to CRY levels in wildtype flies, but not in *cry* mutants (Stanewsky et al. 1998), and CRY light regulation is not affected by TIM or other clock molecules, investigators concluded that CRY served an entrainment function by affecting concentrations of TIM.

Assuming conservation, researchers at first thought that CRY figured in the input pathway in mammals as well, and the initial evidence lent support. Miyamato and Sancar (1998) found that CRY1 and CRY2 occur in retinal ganglion cells as well as the inner nuclear layer of the mouse retina and that CRY1 concentrations oscillate in a circadian manner in the SCN itself. They interpreted this evidence as indicating a photoreceptor role for both CRY proteins. Thresher et al. (1998) offered further support for this view by showing that CRY2 deficient mutants failed to entrain to light. Thresher et al. also observed, however, that CRY2 deficient mutants showed elongated periods in total darkness, a finding that would not be expected if CRY2 figured only in entrainment. Whereas Thresher et al. attempted to explain this finding away, other evidence soon pointed to a different conclusion—that mammalian CRY homologues were not serving the entrainment function but were actually part of the central clock mechanism and not part of the entrainment process. One piece of evidence was the finding that loss of both CRY1 and CRY2 eliminated circadian rhythms in constant darkness and loss of just one resulted in shortened (CRY1) or lengthened (CRY2) periods of oscillation. (van der Horst et al. 1999). Using an assay in cultured cells in which CLOCK-BMAL1 drives a luciferase reporter gene from the *mPer1* gene E-box, Griffin, Staknis, and Weitz (1999) determined that both hCRY1 and hCRY2 produced a specific inhibition of CLOCK-BMAL1 activity. This indicated that the “role for CRYs in the mammalian circadian clock is to inhibit *Per1* gene expression” (p. 769). Citing other evidence that CRY no longer performed an entrainment function, these researchers concluded that CRY1 and CRY2 has supplanted TIM in the mammalian clock. (Figure 2 shows this change as well as the role of each as a dimerization partner with PER.)
The discovery that a conserved component was performing a very different operation in the mammalian clock than it did in *Drosophila* generated a new research question: What performs the entrainment operation in mammals? The clue was again provided by conservation. Melanopsin, a member of the opsin family of photopigments, was discovered in melanophores (melanin pigment containing cells) of the frog *Xenopus laevis* (Provencio et al. 1998). Subsequently, Provencio et al. (2000) reported finding melanopsin in the mammalian inner retina. Linking melanopsin to circadian entrainment required demonstrating its presence in the retinohypothalamic tract that had long been known to provide the input to the SCN. Hannibal had previously identified pituitary adenylate cyclase activating peptide (PACAP) as the neurotransmitter active in the retinohypothalamic tract during subjective day, and he now showed that both melanopsin mRNA and protein are found in the same cells as PACAP (Hannibal 2002). Although knockout of melanopsin alone does not eliminate entrainment, it does reduce it (Ruby et al. 2002), and when the knockout is combined with loss of the rods and cones, entrainment is eliminated (Hattar et al. 2003). Together this provided strong evidence that in circadian entrainment for mammals, melanopsin had assumed the role of CRY.

With melanopsin replacing CRY in entrainment, and CRY replacing TIM in the central oscillator, a natural question is: what happened to TIM? It is still expressed in mammals, but many researchers initially concluded it ceased to perform any clock function (Albrecht 2002). Griffin et al., in the study that suggested CRY and supplanted TIM, had described an antagonistic interaction between either CRY1 or CRY2 and TIM which they took to “suggest cross-regulation among the proteins inhibiting CLOCK-BMAL1 activity within the circadian clock feedback loop” (p. 470). Since deletion of mTim is lethal in embryonic development, the role of TIM could not be settled through a knock-out experiment. The development of the gene knockdown technique in which a reagent complementary to an active gene
oligodeoxynucleotides, αODN) binds to the site, and suppresses its expression, allowed for a more specific examination of the role of mTim in slice preparations. When a αODN targeted to the sequence surrounding the start codon of mTim was applied for prolonged periods, full-length mTIM (mTIM-fl) expression was suppressed. Importantly, so were circadian rhythms (Barnes et al. 2003). When the αODN was only administered for two hours, it resulted in a phase shifts if applied during subjective day (advances when applied early, delays when applied late), but not if administered during subjective night. It was also demonstrated that mTIM-fl interacts with the three mPERs as well as both mCRYs. In the knockdown preparation levels of the mPERs decreased, a finding that corresponds well with the low levels of dPER in dtim mutants and, according to Barnes et al., shows that “mTIM-fl is a functional homolog of dTIM”. Relying on additional evidence, Barnes et al. propose locating “mTim on the negative arm of the molecular feedback loop in the SCN, as it is in the Drosophila clock” (p. 441). In particular, they hypothesized “that mPER2 may be the physiologically preferred partner of mTIM-fl and that this heterodimer could be the functional mammalian counterpart of dTIM:dPER. Addition of mTim to the mammalian clockwork completes a core having each functional homolog of the Drosophila clockwork and emphasizes the highly conserved nature of the biological timekeeping mechanism” (p. 441). At present the question of the role of Tim remains undecided.

5. Conservation as Strategy for Generalizing and Discovery

The initial success in developing a core mechanism for circadian oscillations in Drosophila invited the heuristic of seeking mammalian homologues for the Drosophila genes so as to generalize the explanation to mammals. This proved successful in the case of per where three mammalian homologues were identified, at least two of which play comparable roles in the mammalian clock. Independent work on the mammalian clock resulted in the identification of two additional components of the clock mechanism, Clock and Bmal1, and this prompted the search for homologues in Drosophila. The assumption of conservation and the search for homologues thus serves as a heuristic for generalizing accounts of mechanisms across species.

But it can serve as an even more powerful discovery heuristic when the model does not align perfectly with the target. The quest for the mammalian homolog of cry, which figures centrally in the entrainment mechanism in Drosophila, led to the discovery that in mammals it appears not to figure in entrainment but rather to have usurped the role of tim. Initially this discovery was viewed as showing that Tim had no clock function in mammals, but the fact that it continues to be expressed motivated more detailed studies which have suggested that it might retain a clock function by serving as the preferred dimerization partner of one of the mammalian PER homologues. Discovering a change in the operation CRY performs, however, also prompted a search for what filled the role in playing in Drosophila. This yielded the discovery of a different photopigment, melanopsin, and an investigation into whether it might figure in entrainment in mammals.

Mechanisms are typically investigated in specific model systems chosen because of their accessibility for study. These model systems are often not the primary systems of interest, and the mechanisms in the model systems and those of interest often differ in significant respects. But the assumption of conservation of component parts, operations, and organization undergirds a heuristic for generalizing from studied mechanisms to those of principal interest by seeking
homologues. Moreover, when conservation turns out not to be perfect, it gives rise to other
discovery heuristics. When a component is preserved but appears to be performing a different
operation, investigate what performs its old operation and what happens to the component that
had performed the new operation.

Scientific discovery was once viewed as an intractable problem for philosophers. Several
philosophers who have addressed mechanistic explanation in biology have noted that the search
for mechanisms is often guided by discovery heuristics (Bechtel and Richardson 1993; Darden
and Craver 2002). The conservation of mechanisms further advances this project: Assuming
conservation and searching for homologues is a heuristic for both generalizing explanations and,
when conservation results in variation, for guiding the search for other changes in the
mechanisms.

References

Albrecht, Urs (2002), "Functional Genomics of Sleep and Circadian Rhythm: Invited Review:
  Regulation of mammalian circadian clock genes", Journal of Applied Physiology 92
  (3):1348-1355.
Albrecht, Urs, Zhong Sheng Sun, Gregor Eichele, and Cheng Chi Lee (1997), "A Differential
  Response of Two Putative Mammalian Circadian Regulators, mper1 and mper2, to Light",
  Cell 91 (7):1055-1064.
Barnes, Jessica W., Shelley A. Tischkau, Jeffrey A. Barnes, Jennifer W. Mitchell, Penny W.
  Burgoon, Jason R. Hickok, and Martha U. Gillette (2003), "Requirement of Mammalian
  Cambridge: Cambridge University Press.
Bechtel, William, and Robert C. Richardson (1993), Discovering complexity: Decomposition
  and localization as strategies in scientific research. Princeton, NJ: Princeton University
  Press.
Craver, Carl (2007), Explaining the brain: What a science of the mind-brain could be. New
  York: Oxford University Press.
Darden, Lindley (2006), Reasoning in biological discoveries: Essays on mechanisms, interfield
  relations, and anomaly resolution. Cambridge: Cambridge University Press.
Darden, Lindley, and Carl Craver (2002), "Strategies in the interfield discovery of the
  mechanism of protein synthesis", Studies in History and Philosophy of Biological and
Darlington, Thomas K., Karen Wager-Smith, M. Fernanda Ceriani, David Staknis, Nicholas
  Gekakis, Thomas D. Steeves, Charles J. Weitz, Joseph S. Takahashi, and Steve A. Kay
  (1998), "Closing the circadian loop: CLOCK-induced transcription of its own inhibitors
  per and tim", Science 280 (5369):1599-1603.
Emery, Patrick T., W. V. So, M. Kanedo, J. C. Hall, and Michael Rosbash (1998), "CRY, a
  Drosophila clock and light-regulated cryptochrome, is a major contributor to circadian
Gekakis, Nicholas, Lino Saez, Anne-Marie Delahaye-Brown, Michael P. Myers, Amita Sehgal,
  Michael W. Young, and Charles J. Weitz (1995), "Isolation of timeless by PER Protein


