Circadian regulation of sleep in mammals: Role of the suprachiasmatic nucleus

Ralph E. Mistlberger*

Department of Psychology, Simon Fraser University, 8888 University Drive, Burnaby, Canada BC V5A 1S6

Accepted 7 January 2005
Available online 8 March 2005

Abstract

Despite significant progress in elucidating the molecular basis for circadian oscillations, the neural mechanisms by which the circadian clock organizes daily rhythms of behavioral state in mammals remain poorly understood. The objective of this review is to critically evaluate a conceptual model that views sleep expression as the outcome of opponent processes—a circadian clock-dependent alerting process that opposes sleep during the daily wake period, and a homeostatic process by which sleep drive builds during waking and is dissipated during sleep after circadian alerting declines. This model is based primarily on the evidence that in a diurnal primate, the squirrel monkey (Saimiri sciureus), ablation of the master circadian clock (the suprachiasmatic nucleus; SCN) induces a significant expansion of total daily sleep duration and a reduction in sleep latency in the dark. According to this model, the circadian clock actively promotes wake but only passively gates sleep; thus, loss of circadian clock alerting by SCN ablation impairs the ability to sustain wakefulness and causes sleep to expand. For comparison, two additional conceptual models are described, one in which the circadian clock actively promotes sleep but not wake, and a third in which the circadian clock actively promotes both sleep and wake, at different circadian phases. Sleep in intact and SCN-damaged rodents and humans is first reviewed, to determine how well the data fit these conceptual models. Neuroanatomical and neurophysiological studies are then reviewed, to examine the evidence for direct and indirect interactions between the SCN circadian clock and sleep–wake circuits. Finally, sleep in SCN-ablated squirrel monkeys is re-examined, to consider its compatibility with alternative models of circadian regulation of sleep. In aggregate, the behavioral and neurobiological evidence suggests that in rodents and humans, the circadian clock actively promotes both wake and sleep, at different phases of the circadian cycle. The hypersomnia of SCN-ablated squirrel monkeys is unique in magnitude, but is not incompatible with a role for the SCN pacemaker in actively promoting sleep.

© 2005 Elsevier B.V. All rights reserved.

Theme: Neural basis of behavior
Topic: Biological rhythms and sleep

Keywords: Rat; Mouse; Squirrel monkey; Preoptic area; Dorsomedial hypothalamus; Hypocretin; Subparaventricular zone; Circadian clock gene; EEG; Slow wave sleep

Contents

1. Introduction ........................................................... 430
2. Total daily sleep time in the absence of circadian rhythms: nocturnal rodents .................................................. 432
   2.1. SCN ablation does not affect total daily sleep time in nocturnal rats .................................................. 432
   2.2. SCN ablation may increase total daily sleep time in some strains of mice ........................................... 433
   2.3. Reversible arrhythmicity, by some but not other methods, increases total daily sleep time in some, but not other species. .......................................................... 433
1. Introduction

The modern fields of chronobiology and sleep research, despite separate origins in the late 1950s, have shared for many years a common interest in how the daily (circadian) rhythm of behavioral states is controlled. Given its relevance to human health, well-being, and performance, this is an important problem to solve and remains an active area of research. Over this time, a master circadian pacemaker critical for circadian organization of sleep–wake states has been localized to the suprachiasmatic nucleus (SCN), a retinorecipient cluster of ~16,000 neurons bilaterally distributed at the base of the third ventricle in the anterior hypothalamus [120]. Parallel work has identified an anatomically distributed system of interconnected but neurochemically distinct cell groups specialized for the induction and maintenance of arousal, rapid-eye-movement sleep (REMS) and/or non-REMS (NREMS) [108,109,203,217].

2.5. Daily sleep time and arrhythmicity: summary
2.4. Sleep in circadian clock gene mutants: brave new world
2.3. Insights from sleep deprivation studies
2.2. The SCN may promote sleep intensity in nocturnal rodents
2.1. The SCN alternately opposes and promotes sleep in nocturnal rats
3. Hypersomnia in SCN-ablated squirrel monkeys, revisited
4. Opponent processes: Staying clear on the concept
5. Conclusions

Models 1 and 2 propose a unidirectional or discontinuous effect of the circadian clock on sleep and wake (active modulation of sleep and/or wake circuits during only one portion of the sleep–wake cycle), whereas Model 3 proposes a bidirectional or continuous influence (active modulation of sleep–wake circuits at opposite phases or through all phases of the cycle).

In 1993, Edgar and colleagues proposed an ‘Opponent Process’ model of sleep regulation [69]. This conceptual model is similar to the quantitative ‘Two Process’ model, articulated a decade earlier by Borbely, Daan, Beersma, and colleagues [22,42], to the extent that it attributes sleep timing to the combined influence of the circadian clock and a homeostatic sleep process. In the Two Process model, the circadian influence (so-called ‘process C’) is imparted by modulation of upper and lower thresholds for sleep onset and termination, respectively. Sleep is initiated when a sleep–wake dependent neurophysiological factor (so-called ‘process S’, corresponding to sleep need) reaches the upper
threshold, and sleep is terminated when this factor declines to the lower threshold. The model successfully simulates many phenomenological aspects of sleep but the neural bases of the ‘C’ and ‘S’ factors and the upper and lower thresholds remain to be determined. At a formal level, the Two Process model may be compatible with either unidirectional (our Models 1 and 2) or bidirectional (Model 3) control of the sleep–wake thresholds by the circadian clock. The Opponent Process model is a version of a two process model in which the formal basis of the circadian influence is explicitly specified. According to the Opponent Process model, the circadian clock actively promotes waking, but does not actively promote sleep, i.e., it corresponds to Model 1 in our simplified formulation. This idea was suggested by the observation that a group of 5 squirrel monkeys (*Saimiri sciureus*) sustaining complete ablation of the SCN circadian clock not only lost circadian organization of sleep–wake, but also slept on average ~3.9 h more per 24 h than did 5 intact monkeys, an increase of ~46% [69]. One way that this could occur is if the SCN provided a daily alerting signal that stimulated activity in brain circuits supporting cortical, autonomic, and behavioral arousal, thereby opposing the expression of sleep drive and creating a consolidated daily wake phase. Without this influence, spontaneous bouts of arousal would become fragmented and total sleep time might be expected to increase.

In recent years, the Opponent Process model has been represented, either implicitly or explicitly, as a fact of sleep–wake regulation (e.g., Refs. [8, 25, 51, 72, 94, 117, 132, 258]). For example, a chapter on circadian rhythms in a leading neuroscience textbook states that “sleep and waking are controlled by two opposing factors, a homeostatic drive for sleep and a circadian arousal stimulus” [161]. The neural mechanisms by which the circadian clock influences the sleep–wake system remain to be fully specified, but progress is being made, and it seems timely to reconsider the merits of opponent processes as an organizing concept for empirical findings and further studies. This paper has three main objectives: (1) To highlight lesion and behavioral evidence consistent with a bidirectional (Model 3) rather than strictly oppositional (Models 1 and 2) influence of the circadian clock on sleep–wake states in nocturnal lab rats and humans, (2) to review neurobiological evidence suggesting that the circadian clock actively influences both sleep and wake circuits in nocturnal and diurnal rodents, and (3) to critically re-examine the concept and empirical basis for the Opponent Process view of sleep in squirrel monkeys.

The strategies available for elucidating the neural basis of circadian control of sleep–wake include the traditional methods of lesion and stimulation (how does removal, inactivation or activation of circadian clock cells affect behavioral state, acutely and chronically?), functional neuroanatomy (how is the clock wired to sleep–wake circuits; what are the functional consequences of these connections across the circadian cycle, as determined by single unit electrophysiology or gene expression?), and new methods for manipulating the circadian clock at the molecular genetic level (how do developmental and conditional gene knockouts or knockins affect behavioral state?). To make inferences about the nature of circadian clock...
control of sleep from such manipulations, careful behavioral and electrophysiological analyses of sleep may be critical. For example, a long-sleep phenotype created by a lesion or gene knockout could reflect loss of a circadian alerting signal that normally opposes sleep, or loss of a circadian sleep signal that normally promotes consolidated, restorative sleep, thus diminishing sleep intensity and causing a compensatory increase in total sleep time to achieve presumed recovery benefits of sleep. There may also be interactions with the environment in which the animals are tested (see Section 6 for discussion). Ambiguities in interpreting changes in total sleep time may only be resolved by converging evidence from experiments designed to probe sleep characteristics (e.g., its duration, continuity and intensity) following behavioral perturbations (e.g., sleep deprivation) or under different environmental conditions (e.g., tethering vs. telemetric recording of biosignals; standard impoverished vs. socially enriched environments). Gene knockout approaches are further challenged by the possibility (likelihood? [88]) of pleiotropy; genes at the core of the mammalian circadian clock are expressed elsewhere in the brain, and may participate in aspects of sleep regulation independent of circadian timing. The sleep–wake states are complex phenomena, requiring coordination of diverse neural systems that subserve behavioral, endocrine, and autonomic outputs. Circadian regulation of these states may be relatively simple, or not, but these introductory methodological considerations should serve to prime the reader for interpretive complexity in the pathway to discovery.

2. Total daily sleep time in the absence of circadian rhythms: nocturnal rodents

The primary empirical basis for the Opponent Process model of sleep regulation is the observation that in squirrel monkeys, SCN ablation not only eliminates sleep–wake circadian rhythms but also substantially increases total daily sleep time. The model thus rests on an assumption that if the circadian clock regulates behavioral state exclusively by promoting arousal during one portion of the circadian cycle, then inactivation of the clock will produce hypersomnia. By this logic, a decrease in sleep time after SCN ablation would presumably signify a sleep-promoting effect of the clock, whereas no change in sleep time would signify a balanced driving or gating effect of the clock on sleep and wake circuits. Conceivably, the circadian clock may actively promote both wake and sleep at different phases, but the relative strengths of these effects may differ within or across species, thereby determining whether SCN ablation results in increased, decreased or unchanged total daily sleep time. Consequently, lesion studies can provide supportive, but not decisive evidence in support of any particular model. However, if daily sleep duration either decreases or remains unchanged after SCN ablation in a particular species, this would constitute strong evidence against the validity of the Opponent Process model (Model 1) for that species. Conversely, an increase or no change in sleep duration would constitute evidence against Model 2, whereas Model 3 is compatible with any lesion result.

Given the importance of the SCN ablation result to the Opponent Process model, a reasonable question to address first is whether this result has been replicated in squirrel monkeys and extended to other species. So far, the squirrel monkey study has not been repeated, and sleep studies of other diurnal species with verified SCN ablation are still awaited. The generality of the squirrel monkey result can thus only be evaluated in nocturnal species, primarily those few rodents–rats, mice, and hamsters–that have been popular in sleep and chronobiology research.

In nocturnal rodents, the circadian organization of behavioral state can be eliminated not only by SCN ablation, but also by manipulations of lighting or of genes that mediate circadian timing. A comparison of how these methods for eliminating rhythmicity affect sleep duration may be informative for understanding clock regulation of sleep, and also for understanding the basis for arrhythmicity. Sleep studies of these arrhythmic models have been conducted in one or more species, but unfortunately, no one species has been evaluated using each of these techniques. In the most extensively studied species, the laboratory rat (Rattus norvegicus), the results of the available studies form a coherent picture of sleep regulation that does not support the Opponent Process model. Studies of other species are comparatively few in number, and more diverse in outcome. The results for these other species leave open the possibilities that (1) there may be significant differences in clock control of sleep–wake even among closely related nocturnal rodent species, and (2) different methods for inducing arrhythmicity are probably not functionally equivalent in their effects on SCN output.

2.1. SCN ablation does not affect total daily sleep time in nocturnal rats

Polygraphic sleep studies indicate that complete ablation of the SCN in Norway rats eliminates sleep–wake rhythms but has little or no effect on total daily sleep time or its component NREM and REM stages. This finding has been replicated in several labs, using standard 12:12 light–dark (LD) cycles and dim constant light (LL) or dark (DD) (e.g., Refs. [10,99,100,155,166,220,241,255]). Only one study has reported a significant, albeit very small (~4%) increase in NREMS after SCN ablation in rats, but this was offset by a ~10% decrease in REMS, resulting in no net change in total sleep time [150]. Whether these effects would have been stable over time is not known. In aggregate, these studies indicate that in the nocturnal Norway rat lacking a SCN, there is little or no change in daily sleep duration.
2.2. SCN ablation may increase total daily sleep time in some strains of mice

Two labs have reported on the sleep of mice following SCN ablation. One lab found no change in NREMS or REMS duration in blind or sighted arrhythmic SCN-ablated BALB/c mice recorded in LD [101], but another lab has now reported a significant (~20%), or nearly 2 h/day) increase in NREMS, with no change in REMS, in SCN-ablated C57B1/6j mice recorded in LD and DD [67]. Two caveats are worth noting. First, the daily amount and circadian distribution of sleep in mice exhibit significant variability across strains and laboratories (e.g., Ref. [76,213]), thus it remains to be established whether this effect of SCN ablation generalizes to other mouse strains. Second, by comparison with rats, mice show much greater sensitivity to effects of recording conditions; e.g., even small changes in the weight and flexibility of recording cables can significantly increase or decrease daily sleep amounts in some strains [235], and simply providing a running wheel reduces total daily sleep time (both NREMS and REMS) by ~10% [68] to 14% [248] in C57B1/10j mice. SCN lesions reduce activity levels in nocturnal rodents, and this change alone could account for much if not all of the reported increase in sleep. This may be particularly pertinent to studies of mice, as they are more active than are rats in standard laboratory shoebox cages, with or without wheels. Consequently, interactions between SCN ablation and the recording conditions will need to be carefully examined to interpret any changes in sleep time observed in some SCN-ablated mouse strains.

2.3. Reversible arrhythmicity, by some but not other methods, increases total daily sleep time in some, but not other species

Circadian rhythms of sleep–wake in one or more nocturnal rodent species can be reversibly eliminated by manipulations of light exposure. Rats are particularly susceptible to damping of circadian organization by long-term exposure to LL. Although LL tonically suppresses locomotor activity in nocturnal rats [5], total daily sleep time does not appear to change as the circadian sleep–wake cycle gradually damps out (unpublished analysis of data from Refs. [65,66]). This is a particularly powerful demonstration, as daily sleep amounts were obtained by continuous, within-subject polygraphic recordings as circadian organization ‘unthreaded’ into an entirely ultradian pattern. A more recent study further showed that neither NREMS nor REMS amounts were altered in rats with damped circadian rhythms after 12 days in LL [103].

In another nocturnal rodent, the Djungarian (also called Siberian) hamster (Phodopus sungorus sungorus), circadian sleep–wake rhythms can be rapidly and reversibly eliminated by exposure to a short photoperiod (8 h light/day; [47]), by a 5-h phase delay of the LD cycle [195,196], or by appropriately timed light pulses at night [223]. In the arrhythmic state induced by a short photoperiod, the daily amounts of NREM and REMS are unaltered [48]. In the arrhythmic state induced by a 5-h LD delay shift, daily NREMS increases by ~7%, REMS by ~30%, and total sleep time by ~1.5 h/day [131]. The 24-h profile of behavioral state in the arrhythmic hamsters looks remarkably like a continuation of the normal sleep phase of the circadian cycle (the ‘subjective day’ in nocturnal animals). Sleep studies of arrhythmicity induced by light pulses at night [223] have not yet been conducted.

The differential effects of photic manipulations on sleep duration in Djungarian hamsters suggests that arrhythmicity induced by different means may have a different neural bases. In Djungarians lacking sleep–wake circadian rhythms during exposure to short photoperiods, at least some other circadian rhythms appear to persist [47,48,93,197]. In Djungarians that are arrhythmic following LD shifts, loss of circadian organization extends to physiological variables [195]. Conceivably, short photoperiods may affect circadian organization of behavioral state downstream from the circadian clock, and thus may only mask pacemaker phase (as assessed behaviorally), whereas a 5-h phase delay of LD (involving critically timed light exposure early in the night?) may stop the Djungarian circadian clock at a phase corresponding to subjective day, thereby generating a continuous circadian drive for sleep. Predictions from the latter hypothesis include (1) sleep bouts should be long (rather than fragmented, as in SCN-ablated animals) throughout the 24-h day, comparable to the normal sleep period in rhythmic hamsters (they are; [131]), (2) the arrhythmic hamsters should be more difficult to keep awake during sleep deprivation, comparable to rhythmic hamsters in their subjective day (they are; [131]). (3) SCN ablation should normalize total daily sleep time in arrhythmic animals (not yet tested), and (4) molecular markers of circadian phase should also indicate subjective day (but they do not; photic induction of cFos and Per1 gene expression in SCN neurons is normally limited to the subjective night in rhythmic hamsters, but arrhythmic Djungarian hamsters exhibit high levels of induction in response to light at any time of day; [11]). Although the molecular evidence appears to weigh against the idea that the SCN in arrhythmic Djungarian hamsters is ‘stuck’ in the subjective day, the SCN lesion experiment remains of interest to conduct, given the possibility, considered next, that under certain conditions photic induction of gene expression in the SCN may not mark circadian phase.

2.4. Sleep in circadian clock gene mutants: brave new world

New models for examining sleep regulation in the absence of circadian timing have been generated by selective knockouts or mutations of one or more of the so-called ‘circadian clock genes’ that constitute the molecular gears of the SCN pacemaker [125,213]. The circadian clock is based on interlocking, positive and negative, autoregula-
ory, transcription—post-translation feedback loops between these clock genes and their protein products [92]. Two genes, Clock and Bmal1, encode transcription factors (CLOCK and BMAL1) that form heterodimers and drive expression of two period genes (Per1 and 2) and two cryptochrome genes (Cry1 and 2). The PER and CRY proteins form complexes that feed back to inhibit CLOCK:BMAL1. The clock can be ‘broken’ by double knockouts of either Cry1 and 2 or Per1 and 2; mice of either genotype are arrhythmic in DD [176]. Notably, like Djungarian hamsters driven to arrhythmicity by a LD cycle shift, photic induction of Per1 and/or Per2 persists in arrhythmic Cry1,2 null mice [176]. Thus, photic induction of Per1 can be dissociated from the subjective night and can occur in the absence of circadian cycling. Conceivably, SCN clock cells in arrhythmic Djungarian hamsters may be in a cellular state comparable to a Cry1,2 null mouse, and this state may be associated with cellular outputs (action potentials or release of diffusible factors; see below) that on balance promote sleep. If so, then Cry1,2 null mice should also be long sleepers, and they are; the one available study reported ~18% more NREMS per 24-h day, with consolidated as opposed to fragmented sleep bouts [253].

These observations raise the possibility that some genetic and environmental manipulations can move circadian pacemaker neurons to a unique cellular state associated with characteristics of both the subjective night (e.g., photic induction of Per genes) and the subjective day (e.g., outputs continuously promoting sleep). Hypersomnia associated with arrhythmicity in animals with an arrhythmic SCN, and in many peripheral tissues outside of the brain [92]. Changes in sleep duration following clock-gene mutations could therefore be entirely unrelated to normal SCN function. More complex still is the possibility that the sleep phenotype reflects an interaction between effects of gene mutations both within and outside of the SCN clockworks.

2.5. Daily sleep time and arrhythmicity: summary

The lab rat data tell a consistent story; contrary to predictions of the Opponent Process model, arrhythmicity following SCN ablation or long term exposure to LL has little or no effect on total daily sleep time. The results of mouse SCN ablation and gene knockout studies are more diverse, as are the Djungarian hamster light manipulation data. It seems likely that loss of circadian rhythmicity by SCN removal, photic manipulation, or altered genes are not equivalent models for analysis of the nature of SCN outputs. More work is needed to fill gaps of knowledge, as none of the few species studied so far have been screened through each of these models for inducing arrhythmicity. There is also a lack of data on SCN ablation, arrhythmicity, and sleep in diurnal animals other than the squirrel monkey, and such studies are awaited with interest. More troubling is the interpretive ambiguity; changes in total daily sleep time in arrhythmic animals may have multiple causes, and appear to be compatible with more than one model of circadian control of sleep. Important insights, however, can be obtained by studies of normal and arrhythmic animals in which sleep regulatory mechanisms are challenged by sleep deprivation.

3. Insights from sleep deprivation studies

3.1. The SCN alternately opposes and promotes sleep in nocturnal rats

Several studies have examined recovery sleep after short-term total sleep deprivation in nocturnal lab rats with SCN lesions, to characterize sleep homeostasis in the absence of circadian timing [155,238,241]. These studies show that SCN ablation does not disrupt sleep homeostasis. Intact and SCN-ablated rats exhibit a similarly modest increase in total daily sleep time over the first few days of recovery (e.g., during 3 days of recovery sleep, SCN-ablated and intact rats recouped ~30% of sleep ‘lost’ during a 24 h sleep deprivation; [155]) but an immediate and strong increase in sleep intensity, operationally defined by changes in cortical EEG during NREMS, including total power, and the amplitude and incidence of waves in the ‘delta’ frequency band (~0.5–4.0 Hz, also referred to as slow-waves or slow-wave activity, SWA). These data confirmed that homeostatic and circadian regulation of sleep are mediated by distinct neural substrates. This was an important but not unexpected observation.
What is more interesting about these sleep deprivation data, although apparently overlooked in the sleep regulation literature, is what they reveal about the means by which the circadian clock imposes daily rhythms on sleep–wake states. A comparison of recovery and baseline sleep in SCN-ablated and intact rats clearly shows that the circadian clock actively promotes wake during the active phase and sleep during the rest phase of the daily sleep–wake cycle, i.e., supporting Model 3 in our simplified formulation. This conclusion follows from the data illustrated in Fig. 2 (adapted from Ref. [155]). Panel A shows the familiar circadian rhythm of sleep in intact rats and the lack of circadian variation in SCN-ablated rats (confirmed to be absent in each subject). Following 24 h of sleep deprivation, recovery sleep in the intact rats is obviously constrained by circadian timing; this is best illustrated by directly comparing the first 12 h of recovery in the intact rats with the baseline data from the SCN-ablated rats (Panel B). Despite 24 h of sleep loss, the intact rats express less sleep during the first 12 h of recovery, which occurred during their usual daily wake phase (lights-on period), than do the SCN-ablated rats that are not sleep deprived. Thus, during the normal wake phase, the circadian clock opposes the expression of recovery sleep, as proposed by the Opponent Process model. But, as illustrated in Panel C, this active influence switches direction during the normal sleep phase. Here, sleep during the first 12 h of recovery in the SCN-ablated rats is contrasted with sleep during the normal baseline sleep period in the intact rats. The intact rats are not sleep deprived, yet they sleep significantly more than the SCN-ablated rats that have been awake for 24 h. The same is true for both NREMS and REMS (data not shown), supporting that version of Model 3. The circadian clock must therefore actively promote sleep. If it did not, then very sleep deprived rats lacking a SCN should sleep at least as much as intact rats during their normal sleep phase, if not more.

A study of selective REMS deprivation in SCN-ablated and intact rats provides additional evidence that the circadian clock actively promotes REMS during the usual

---

**Fig. 2.** (A) Sleep in intact (bold line, closed circles) and SCN-ablated (dashed line, open squares) rats, as a percent of total time in 2-h time blocks for 24 h prior to and 24 h after a complete day of sleep deprivation (see Ref. [96] for methodological details). Intact rats were recorded in LD (lights-on during hours 13–24). (B) Comparison of sleep in the intact rats during the first 12 h of recovery (lights-off, corresponding to 1–12 of the recovery day in panel A) with sleep in the SCN-ablated rats during the first 12 h of the baseline day (corresponding to hours 1–12 of the baseline day in panel A). Cumulative sleep totals over this interval were significantly different between groups by MANOVA. (C) Comparison of sleep in the intact rats during the 12 h lights-on (sleep) period of the baseline day (corresponding to hours 13–24 of the baseline day in panel A) with sleep in the SCN-ablated rats during the first 12 h of the recovery period after sleep deprivation (corresponding to hours 1–12 of the recovery day of panel A). Cumulative sleep totals over this interval were significantly different between groups by MANOVA.
sleep period [255]. Rats were subjected to state-dependent REMS deprivation, which yielded a measure of attempts to initiate REMS. The number of REMS attempts increased significantly during REMS deprivation in intact and SCN-ablated rats, but the greatest increase was evident in intact rats during their subjective day (the sleep phase of the circadian cycle). This is consistent with other studies showing strong circadian regulation of REMS propensity in humans and other species (e.g., Refs. [40,58,142]).

3.2. The SCN may promote sleep intensity in nocturnal rodents

Sleep intensity is a construct based on descriptive evidence that the levels of some electrophysiological and behavioral correlates of sleep vary within sleep and are sensitive to recent wake–sleep history. As noted, a correlate of special interest is EEG SWA during NREMS. In humans and rats, SWA is generally associated with elevated auditory arousal thresholds [171,251], which is one operational definition of sleep depth or intensity. SWA is maximal early in the sleep period, declines exponentially as sleep progresses, is proportional to prior wake duration, is reduced by napping during the usual wake period and is enhanced by short-term (e.g., 24 h) sleep deprivation (e.g., Refs. [24,73,74,156,237,250,241,245]). Suppression of SWA by auditory stimulation during the first 3–5 h of sleep stimulates a strong increase of SWA during the remaining sleep period that is sufficient to recoup normal amounts before spontaneous sleep termination [56,62]. SWA and its rate of decline during sleep are therefore viewed as electrophysiological correlates of sleep intensity and of the sleep recovery process. The time constants describing the buildup and decline of SWA during daytime napping and nocturnal sleep, respectively, have been used for quantitative modeling of sleep homeostasis (e.g., Refs. [3,42,75,76,98]).

If SCN-ablated rats sleep less during the first 12 h of recovery from 24 h sleep deprivation than do non-deprived intact rats during their normal 12 h (lights-on) sleep period, could that be because recovery sleep in SCN-ablated rats, freed from circadian constraints, is more intense, enabling a more concentrated, rapid recovery? The available evidence suggests that if anything, the sleep of SCN-ablated rats is less, not more intense. Two groups have reported that SCN-ablated rats, by comparison with intact rats, exhibit significantly less NREMS SWA, both prior to and following sleep deprivation, measured in the time domain (period-amplitude analysis; [16,156]) and frequency domain (Fast Fourier Transform; [241]). SCN-ablated rats housed individually under standard (i.e., environmentally impoverished) laboratory conditions do not spontaneously stay awake for extended periods; therefore, a low level of SWA during baseline sleep recordings could simply reflect the dependence of SWA on the duration of prior wake (e.g., Ref. [157]). If so, then SWA in SCN-ablated rats should recover to normal levels after 24 h of sleep deprivation. However, it does not. SCN-ablated rats do show a significant increase in SWA during recovery sleep, but the magnitude of change is similar to that exhibited by intact rats, and the absolute levels remain lower even by comparison with the baseline sleep of intact rats (e.g., Refs. [16,156]).

The one study, noted in Section 2.1, that reported a small (4%) but significant increment in NREMS in SCN-ablated rats also reported a 15% increase in a normalized measure of SWA, namely, the ratio of power in the 2–4 Hz band in NREMS to the same in REMS [150]. However, in this normalized form, the results cannot be compared to the other studies, as these reported absolute values. There is no way to know from ratio data alone how SWA in NREMS is altered as a consequence of SCN ablation; SWA could be decreased in all states, and if the decrease were greater in REMS, then NREMS SWA would appear to be enhanced by comparison. Comparisons across studies are also compromised by the use of different frequency bands to define SWA (2–4 Hz in [150]; 0.5–3 Hz in [241]; 1–4 Hz in [16,156]), given evidence of heterogeneity in the genesis of slow-wave sub-bands. EEG slow waves in the 1–4 Hz range are mediated by calcium-dependent low threshold currents and hyperpolarization-induced inward currents in thalamocortical neurons [147,178], whereas very slow waves recurring at <1 Hz are cortically generated [201,226] and, unlike 1–4 Hz waves, do not require functional α-1G T-type calcium channels [134]. To facilitate progress in clarifying the role of the SCN in sleep intensity, greater standardization of methods for quantifying and reporting SWA and other sleep parameters is desirable.

The significance of low (absolute) levels of SWA after SCN ablation in rats is uncertain, but one possibility is that the SCN promotes SWA and sleep intensity independently of its role in extending the duration of spontaneous wake episodes. While homeostatic factors clearly dominate regulation of SWA in intact mammals, the SCN may participate by contributing, via polysynaptic pathways, to the inhibition of ascending reticular activating pathways that is necessary for the expression of thalamocortically generated SWA during NREMS [226]. If the SCN does influence sleep intensity, then a circadian variation of SWA would be predicted. Examples of daily variations in NREMS SWA are common, but most of these are inconclusive due to uncontrolled variation in prior wake time. A few studies, however, do report circadian modulation of SWA independent of prior wake time, consistent with a role for the SCN in sleep intensity (e.g., Refs. [29,30,59,63,122,185,234]).

Alternatively, the reduction in SWA levels after SCN ablation may be unrelated to changes in sleep intensity. Quantitative modeling suggests that the rate of change of SWA within sleep may be a more fundamental correlate of sleep intensity than is the absolute level of SWA [13]. Also, dissociations between SWA and sleep restorative processes...
have been noted. Rats recovering from 24 h of sleep deprivation continue to accumulate excess sleep for at least 3 recovery days, despite a decline of high-amplitude, slow-wave-enriched NREMS to below baseline levels (a so-called ‘negative rebound’; [24,76,77,155,212]) after a strong initial increase. Rats recovering from 4 days or more of total sleep deprivation exhibit a prolonged increase of REMS but no elevation of high EEG amplitude, slow-wave-enriched NREMS [189]. Mice lacking functional α-1G T-type calcium channels fail to generate SWA in NREMS, but exhibit a small (~8%) decrease in total sleep time, rather than an increase as might be expected if sleep were less intense and thus less restorative [133]. These results indicate that the level of SWA and high-amplitude EEG NREMS are likely epiphenomenal to the recovery sleep process, and under some conditions can be uncoupled from homeostatic processes that regulate sleep duration.

To examine the role of the SCN in sleep intensity further, it would be of interest to know whether other putative correlates of sleep intensity, such as arousal threshold, are similarly affected by SCN ablation in rats. It would also be informative to examine sleep intensity measures in intact rats that are driven to arrhythmicity by long-term exposure to LL. Is reduced NREMS SWA related to absence of circadian rhythmicity, or absence of the SCN? Evidence from circadian clock gene knockout animals suggests that loss of circadian organization per se does not necessarily result in reduced SWA; Cry1,2 double knockout mice are arrhythmic in DD, but exhibit increased SWA in NREMS, despite an increase of total daily NREMS [253]. These observations strengthen the notion that SCN outputs at some circadian phases may promote both sleep duration and intensity in nocturnal rodents.

4. Circadian clock regulation of sleep in humans

Analyses of sleep in intact and SCN-ablated Norway rats indicate that the Opponent Process account, as formulated, is incorrect for this species. Regulation of sleep–wake states in rodents may be of intrinsic interest to neurobiologists, but implicit in the support of basic sleep research by publicly funded national health institutes is the assumption that principles revealed from careful study of animal models will be of relevance to understanding human sleep and its disorders. It has been suggested that ‘the polyphasic sleep patterns of nocturnal rodents are so profoundly different from the consolidated sleep–wake patterns of diurnal primates that one must critically question sweeping generalizations based only on laboratory rat data’ [61]. This statement raises an important question; is circadian control of sleep in humans more similar to rodents or to squirrel monkeys? Over the past 30 years, studies of naps (e.g., multiple sleep latency tests), short and ultrashort sleep–wake schedules, short-term constant routines, long-term temporal isolation, and forced desynchrony have produced a coherent body of evidence consistent with the view that the circadian clock in humans plays a bidirectional role in sleep–wake regulation, alerting at some phases, and promoting sleep at other phases (for reviews, see Refs. [40,107]). These studies clearly demonstrate that the onset and duration of the main daily sleep episode and of REMS are strongly controlled by circadian phase. The wake-promoting and sleep-promoting functions of the circadian clock are further revealed by a comparison of the daily sleep–wake rhythm and the rhythm of sleep propensity, as measured by sleep latency, the ability to resist sleep, and indices of cognitive alertness; in humans entrained to local time, sleep propensity is lowest in the evening, near the usual onset of sleep and the presumed peak of homeostatic sleep drive, and highest in the early morning, when body temperature reaches its daily minimum during the last hours of sleep, at the presumed nadir of homeostatic sleep drive. This ‘paradoxical’ phase relationship between the usual times of sleep and wake onset and the circadian rhythm of sleep propensity has been hypothesized to explain how sleep in humans is consolidated into a single major episode per circadian cycle; sleep early in the night is facilitated by a high sleep debt load, while sleep late in the night is facilitated by maximal circadian inhibition of arousal or excitation of sleep [57].

Note, however, that the paradoxical phase relation between sleep onset and sleep propensity largely disappears in subjects studied in temporal isolation; under these conditions, humans exhibit a free-running sleep–wake cycle, with a self-selected sleep onset time much closer to the body temperature minimum when sleep propensity is maximal (e.g., Ref. [7]). Thus, while paradoxical phasing may contribute to sleep consolidation, it evidently is not essential.

According to the Opponent Process concept, the circadian clock need not actively drive sleep (e.g., by facilitation of sleep-on neurons, or inhibition of wake-on neurons) to produce the daily waveform of sleep propensity, whatever its phase relation to self-selected sleep onset time. Rather, in entrained humans, the circadian clock could actively drive arousal in opposition to sleep during the day, and this drive could gradually wane through the early night, until it is entirely absent near the body temperature minimum when sleep propensity peaks. Thus, the circadian clock could play a passive gating role rather than active driving role in the expression of maximal daily sleep propensity late in the sleep period. From the behavioral evidence alone, there is no way to rule out either the passive gating or active driving models of these data (corresponding to Models 1 and 3, respectively). Indeed, the human sleep propensity data could also be explained by assuming that the circadian clock actively drives sleep-on circuits, and passively gates the expression of the daily minimum of sleep propensity late in the wake period (Model 2). The Opponent Process account of human sleep regulation would be favored only by virtue of convergent evidence from other approaches, e.g., evidence that there is a selective fragmentation and
reduction in daily wake time following selective disruption of circadian clock function in humans, as in SCN-ablated squirrel monkeys.

One source of such evidence is from studies of sleep in humans with brain damage (reviewed in Ref. [121]). It has long been recognized that injuries to the hypothalamus can grossly disturb sleep–wake regulation in humans. Von Economo [243] described sleep–wake disorders caused by encephalitis; postmortem inspections revealed that insomnia and hypersomnia were associated with damage to the anterior and posterior hypothalamus, respectively. However, these lesions can in no way be described as anatomically specific to the SCN, and it is the anterior lesions that would be more likely to involve the SCN (which produce insomnia, not hypersomnia as predicted by the Opponent Process model). A more recent case study describes a profound sleep–wake disturbance caused by removal of a tumor enveloping the optic chiasm [36]. Damage in the ‘suprachiasmatic region’, including the pituitary stalk (which was absent), was confirmed by MRI. Although cited as evidence for the Opponent Process model [61], this case study did not report increased total daily sleep time. Rather, the most prominent symptoms were bouts of hypersomnolence alternating with alertness and extreme excitement during the day, and repeated awakenings from sleep during the night (15 full awakenings in one overnight polysomnographic assessment). The presentation is thus overwhelmingly one of a failure to consolidate both wake and sleep, and not a selective daytime hypersomnolence. Loss of circadian organization was further evident in a continuous 2-week behavioral observation. This supports an essential role for the SCN in regulation of circadian rhythms in humans, but suggests that, if anything, the SCN actively promotes both wake and sleep. African trypanosomiasis infection is another condition associated with fragmentation of both sleep and wake. Without hypersomnia [139].

Also cited as support for the Opponent Process model is the relationship between sleep and circadian rhythms early in life. Human infants sleep a lot, and initially have weak or no circadian organization [154,194]. Over the first months of life, sleep duration declines as circadian organization emerges. However, this correlation cannot be taken to infer that the reduction of total daily sleep time is caused by the maturation of a wake-promoting signal from the circadian clock. These may well be independent developmental processes. Indeed, during advanced age, circadian organization and sleep consolidation weaken, and there is more daytime napping and less nocturnal sleep. At this stage in life, there is, if anything, less, not more total daily sleep; the correlation reverses sign. Perhaps the most common sleep complaint of the aged is reduced sleep efficiency and unwanted early morning awakening, i.e., repeated nocturnal arousals and an inability to sustain sleep throughout the night [19,60]. These changes have been interpreted as reflecting an age-related decline in active circadian drive for sleep during the subjective night [60].

These results are consistent with the view that the circadian clock in humans, like rodents, actively promotes both wake and sleep at different phases through its cycle. There is nothing in human sleep phenomenology that is uniquely accounted for by an exclusively opponent process mechanism (Model 1). The fact that humans tend to be monophasic (possibly biphasic) sleepers (i.e., 1 or 2 main bouts of sleep/circadian cycle), while rats are polyphasic sleepers (many bouts, but a preponderance of sleep in the day) may belie a more fundamental formal similarity of circadian regulation in these species. It is worth noting that when the recording conditions used in human and rodent studies are made more similar, the observed sleep patterns also become more similar. For example, humans confined to bedrest for several days (analogous to a rodent confined to a standard, small recording box?) exhibit much more fragmented (polyphasic) sleep [29], whereas laboratory mice with free access to a running wheel exhibit more consolidated sleep and wake bouts than when the wheel is locked [249]. No doubt, mice and other rodents in their natural habitats remain substantially polyphasic sleepers, but the point to underscore is that differences between rats and humans in sleep regulation may be more superficial than profound.

This should be good news to sleep researchers, given that neurobiological and genomic investigations of sleep–wake mechanisms are conducted primarily using rodents. Whether the formal similarity of circadian regulation of sleep in man and rodent extends to the neural systems level remains an empirical question.

5. Circadian clock outputs to sleep–wake circuits

A general prediction that follows from analysis of rat and human behavioral data is that the circadian clock will be found to exert a continuous action on sleep–wake circuits throughout its circadian cycle, or, minimally, at the peak and trough of the sleep propensity curve, and that its influence will not be limited to facilitation of arousal during the daily wake period, as proposed to explain hypersomnia in SCN-ablated squirrel monkeys. In other words, during the sleep phase, the clock is not functionally absent. If it were functionally absent, then intact rats during the sleep phase would not sleep significantly more than SCN-ablated rats recovering from 24 h of sleep deprivation, and it seems unlikely that humans would show the paradoxical phasing of sleep onset and sleep propensity rhythms under entrained conditions, the inability to sustain either extended wake or extended sleep after SCN-area damage, and the progressive difficulty sustaining sleep throughout the night in old age. Ultimately, the ‘proof’ of conceptual models will be in the neurobiological analysis. This section will therefore consider neurobiological evidence relevant to these conceptual models. Of special interest are data addressing the following questions: (1) Does the SCN send output signals during both
the subjective day and the subjective night, or is its output restricted to only one phase of the circadian sleep–wake cycle? (2) Does the SCN have monosynaptic or strong disynaptic connections to sleep-active and wake-active cell groups known to regulate sleep and wake, respectively, or does it project to only one side of the sleep–wake neural system? (3) Is there any evidence that nocturnal and diurnal animals differ in these respects?

5.1. Circadian profile of SCN neural activity

The SCN contains a population of cell-autonomous circadian oscillators that express circadian rhythms of output in the form of action potentials and multiple signaling molecules [96,215,249]. Clearly, the daily phasing of SCN rhythmicity must inform any discussion of how the SCN drives circadian rhythms in other neural systems. Integrated measures of SCN neural and metabolic activity vary approximately sinusoidally over the circadian cycle, with peak values during the middle of the subjective day in nocturnal and diurnal animals [86,204,206,207,209]. However, it would surely be simplistic to argue from this that SCN effects on behavioral state are limited to the subjective day. The SCN is not silent at any phase of its cycle and single units and small neural ensembles exhibit considerable heterogeneity in the precise phasing of their daily peaks of action potentials. Some multiple or single units peak in antiphase to the SCN average while others may be phase displaced by ~12 h [96,105,167,206,207]. Vasoactive intestinal polypeptide and vasopressin release from SCN neurons in vitro exhibit circadian rhythms with peaks 6 h apart [215]. SCN output neurons characterized by projections to the supraoptic nucleus or arcuate nucleus exhibit maximal discharge at dawn and/or dusk [198]. Differential phasing of gene expression rhythms across (or defining) subregions of the SCN has also been detected; e.g., a subset of cells in the dorsal, central SCN in mice exhibits peak Per1 and 2 expression in antiphase with the expression rhythm in SCN vasopressin neurons [9,118]. Differential, topographically organized circadian phasing of Per1 gene expression has now been demonstrated at the single neuron level by real-time imaging of SCN slices or organotypic cultures from transgenic mice carrying mPer1-promoter driven luciferase [256] or green fluorescent protein [186] reporter genes. Phase heterogeneity may reflect independent (albeit presumably coupled) endogenously oscillating SCN compartments, gating (e.g., by disinhibition) of one compartment by another endogenously oscillating compartment, or (where observed in vivo) driving by one or more topographically organized rhythmic inputs from other oscillators outside of the SCN (e.g., the retina; [133]).

In addition to rates of spiking, patterns of SCN neural activity may also vary from subjective day to night (e.g., Ref. [198]). Rate and pattern of spiking may determine the ratios of small molecule to peptide neurotransmitters co-released from SCN terminals at different circadian phases. Changes in ratios of co-released signaling molecules could significantly alter the functional effects of SCN outputs to effectors for behavioral state, conceivably shifting the weight of these between primarily excitatory and inhibitory at different circadian phases (e.g., Refs. [199,200]). Differential phasing of SCN unit activity and patterning of co-released transmitters over the circadian cycle may well be important for understanding how SCN outputs organize sleep–wake states.

5.2. SCN efferents: overview

SCN efferents have been mapped in some detail in several mammalian species (e.g., Refs. [2,27,43,54,129,165,225,244]). These projection maps show a high degree of conformity across studies and species. Although the bulk of SCN projections are intra-hypothalamic, even a cursory examination of SCN terminal fields reveals that SCN clock cells are within one or two synapses of most of the major sleep–wake regulating areas. These areas form a widely distributed but interconnected neural system, with specific neurobehavioral correlates of behavioral states controlled by cell groups in the medulla, pons, midbrain, posterior and anterior hypothalamus, preoptic area, basal forebrain, or thalamus [108,109,203,217,230]. The hypothalamus and preoptic areas in particular play a critical role in actively promoting both sleep and wake. The clinical observations of von Economo [243] and experimental animal work of Ranson [188] and Nauta [169] first drew attention to the anterior hypothalamus as the site of an active sleep mechanism and the posterior hypothalamus as the site of a wake-mechanism. Recent studies have substantiated and greatly extended the evidence for interacting sleep–wake circuits in these areas. As the following review shows, there is already considerable neuroanatomical and neurophysiological evidence that the SCN communicates with preoptic and hypothalamic sleep–wake areas both directly and indirectly, via classical synapses and/or paracrine secretion. Lesion studies are also reviewed to determine which if any of these projections has a special role in circadian control of sleep–wake. Knowledge gaps and methodological limitations of the available studies are identified. Due to the lack of data on diurnal species, this review must by default emphasize nocturnal animals.

5.3. SCN efferents to the subparaventricular zone are critical for sleep–wake rhythm

The major efferent pathway from the SCN extends just a few millimeters or less to the subparaventricular zone
(sPVZ), which extends dorsally from the SCN to the ventral border of the hypothalamic paraventricular nucleus (PVN) [244] (Fig. 1). The sPVZ in turn projects to all of the other major SCN terminal fields, suggesting a role for these neurons in modulating circadian signals. Excitotoxic, axon-sparing lesions of the ventral region of the sPVZ, which do not damage SCN neurons, disrupt sleep–wake and activity rhythms, but have less impact on circadian organization of body temperature, thus the ventral sPVZ appears to be an important, specialized route through which circadian signals reach sleep–wake executive circuits [138,162]. Like SCN ablation, sPVZ lesions do not affect daily amounts of either NREMS or REMS in rats [138].

5.4. SCN projections to rostral ‘sleep’ areas

NREMS-active neurons in the mammalian brain have been found in the caudal medulla and the preoptic/anterior hypothalamic area just rostral to the SCN [108]. The ventrolateral preoptic area (VLPO) has received particular attention in recent years for an executive role in sleep regulation [233]. VLPO neurons exhibit increased firing rates [232,233] and cFos expression [214] during sleep, and VLPO lesions induce insomnia [137,231]. The SCN projects to VLPO neurons both directly and indirectly [34,55,173,228]. The direct projection is considered sparse, and appears to include both excitatory (via glutamate) and inhibitory (via GABA) inputs ([228]; for additional evidence of glutamatergic SCN efferents to the preoptic area, see Ref. [124]). Significant indirect projections are via the sPVZ, dorsomedial hypothalamus (DMH), lateral hypothalamus (LH), medial preoptic area (MPOA), and a few other areas [34,55,228]. Excitotoxic lesions eliminating 80–90% of VLPO neurons reduced daily NREM sleep by half but only modestly attenuated the circadian sleep–wake rhythm in rats tested in LD [137]. However, the effects of total VLPO lesions in rats recorded in constant dark (DD) have not yet been reported. Given the strong direct and indirect inputs from the SCN to VLPO neurons both directly and indirectly [232,233], and cFos expression [214] during sleep, and VLPO lesions induce insomnia [137,231]. The SCN projects to VLPO neurons both directly and indirectly [34,55,173,228]. The direct projection is considered sparse, and appears to include both excitatory (via glutamate) and inhibitory (via GABA) inputs ([228]; for additional evidence of glutamatergic SCN efferents to the preoptic area, see Ref. [124]). Significant indirect projections are via the sPVZ, dorsomedial hypothalamus (DMH), lateral hypothalamus (LH), medial preoptic area (MPOA), and a few other areas [34,55,228]. Excitotoxic lesions eliminating 80–90% of VLPO neurons reduced daily NREM sleep by half but only modestly attenuated the circadian sleep–wake rhythm in rats tested in LD [137]. However, the effects of total VLPO lesions in rats recorded in constant dark (DD) have not yet been reported. Given the strong direct and indirect inputs from the SCN to VLPO neurons both directly and indirectly [232,233], and cFos expression [214] during sleep, and VLPO lesions induce insomnia [137,231]. The SCN projects to VLPO neurons both directly and indirectly [34,55,173,228]. The direct projection is considered sparse, and appears to include both excitatory (via glutamate) and inhibitory (via GABA) inputs ([228]; for additional evidence of glutamatergic SCN efferents to the preoptic area, see Ref. [124]). Significant indirect projections are via the sPVZ, dorsomedial hypothalamus (DMH), lateral hypothalamus (LH), medial preoptic area (MPOA), and a few other areas [34,55,228].

The SCN, sPVZ, and DMH project in parallel to a population of neurons containing the neuropeptides hypocretin-1 and -2 (HCRT-1 and HCRT-2, [50]; also known as orexin-A and -B, [202]). Converging evidence indicates that HCRT neurons are an important component of the arousal system. HCRT neurons are distributed widely in the perifornical, dorsolateral, and posterior hypothalamic areas,

the extent of this role remains to be clarified. Of first-order importance will be determining: (1) whether confirmed ablations of sleep-active neurons in these areas affect circadian organization in DD, (2) whether SCN projections (direct and indirect) are to sleep-active neurons, or to other cell types (e.g., wake-active) found in these areas, and (3) whether SCN inputs to particular cell types are primarily excitatory, inhibitory, or alternating with circadian phase.

5.5. SCN projections to caudal ‘wake’ areas

The SCN also has monosynaptic and polysynaptic connections to hypothalamic nuclei implicated in arousal, including the DMH, LH, and posterior hypothalamic (PH) [2,34,54,244]. Again, the monosynaptic pathways are considered light by comparison with polysynaptic pathways involving the ventral sPVZ [35,55,228]. In rats, electrolytic or excitotoxic lesions of the DMH attenuate or eliminate circadian rhythms of sleep–wake, feeding, locomotor activity, corticosterone/cortisol secretion, and neural activity in the locus coeruleus (LC, the site of noradrenergic neurons important for arousal and attention [17]), while largely sparing circadian rhythms of body temperature and plasma melatonin [8,14,35,193]. Excitotoxic lesions also modestly decrease total daily wake time (~7.5%; [35]). Notably, circadian organization of sleep–wake was assessed in DD, ruling out LD masking as an explanation for the residual low amplitude circadian rhythms of NREMS and wake evident in rats with DMH ablation (see Figs. 4a,b in Ref. [35]). Thus, the DMH appears to be an important link in circadian timing of sleep–wake and wake-related functions, possibly serving to integrate circadian wake-promoting signals from the SCN and sPVZ with other neural and endocrine inputs, and to distribute these signals to hypothalamic and extrahypothalamic areas mediating specific behavioral and endocrine correlates of waking. The DMH also projects to the sleep-active POA, providing another potential avenue for active control of sleep by the SCN [34,53–55].

The interdependency of circadian rhythms affected by DMH lesions does raise an unresolved interpretive issue; is disruption of sleep–wake rhythms secondary to loss of circadian organization of feeding, locomotion, and/or endocrine variables? These effects may be independent, as one study of DMH-ablated rats reported persistence of circadian locomotor activity rhythms despite a complete loss of circadian feeding rhythms [193]. Nonetheless, any procedure that eliminates the feeding rhythm is likely to at least partially attenuate the daily rhythm of sleep–wake.

The SCN, sPVZ, and DMH project in parallel to a population of neurons containing the neuropeptides hypocretin-1 and -2 (HCRT-1 and HCRT-2, [50]; also known as orexin-A and -B, [202]). Converging evidence indicates that HCRT neurons are an important component of the arousal system. HCRT neurons are distributed widely in the perifornical, dorsolateral, and posterior hypothalamic areas,
and send excitatory outputs to the entire monoaminergic arousal system, including histamine neurons of the tuberomammillary nucleus (TMN), serotonin neurons of the dorsal and median raphe, noradrenaline neurons of the LC, and acetylcholine neurons of the pedunculopontine tegmentum and basal forebrain [182,203,218]. Intracerebral microinjections of HCRTs during the usual sleep period induce arousal and activity [70,89,102,110], at least partly via actions on TMN histamine neurons [97]. Activation of HCRT neurons is correlated with spontaneous or induced arousal and/or locomotion in several species [71,119,205,239,246,254]. Notably, HCRT cell lesions, prepro-HCRT knockouts or HCRT receptor knockouts in mice, rats, dogs, and/or humans cause symptoms of narcolepsy, including fragmentation of sleep and wake episodes, attacks of hypersomnia and cataplexy during the usual daily wake period, and sleep-onset REMS episodes [18,32,81,91,183,236,252]. HCRTs clearly play a critical role in the regulation of behavioral states.

One hypothesis about the nature of this role is that HCRT neurons mediate circadian clock-dependent alerting [259]. Consistent with this idea, HCRT levels in the cerebrospinal fluid (CSF) of squirrel monkeys exhibit a daily rhythm with a peak late in the daily wake period, when clock-dependent alerting is assumed to be maximal [259]. Rats also exhibit peak HCRT levels in plasma [258] and CSF [49,52] during the latter portion of the daily wake period. In orexin/ataxin-3 transgenic rats, postnatal degeneration of all HCRT neurons is associated with significant, albeit modest, attenuation of waking at the beginning and end of the daily wake period [18].

Other work, however, indicates that HCRT neurons are neither essential for circadian organization of sleep–wake, nor strictly a ‘hand’ of the circadian clock. In the genetic lesion and knockout models, the daily rhythms of wake and NREMS are largely preserved in both LD (e.g., Refs. [4,18,32,91,153,160,252]) and DD [160]. Although the daily rhythm of REMS has been described as either abolished or inverted [18,32,81,91,252], this likely reflects a failure to discriminate between cataplexy and true REMS; when these states are scored separately in HCRT-KO mice, cataplexy is found to be confined to the normal wake period, and the daily rhythm of REMS is essentially normal [160]. Developmental compensation in HCRT-KO mice remains a concern; thus, it will be important to confirm this latter result using the orexin/ataxin-3 ablation model. Also of note, HCRT-KO mice exhibit fragmentation of both wake and sleep, indicating that the defect is not a selective failure to consolidate wake [160].

Unlike the various genetic lesion models, ablation of HCRT neurons by intracerebral injections of the ribosome-inactivating protein saporin conjugated to HCRT-2 does significantly attenuate daily rhythms of wake, NREMS and REMS in adult rats in LD [82,81]. However, this is a much less selective lesion, as the toxin kills any cell expressing the receptor for HCRT-2. Consequently, results from this model alone cannot be used to infer a special role for HCRT neurons in circadian organization of behavioral state.

Correlational studies further weaken any hypothesis that the activity of HCRT neurons represents the phase of the SCN circadian pacemaker. HCRT neurons are active in association with behavioral arousal at any circadian phase; e.g., in squirrel monkeys, CSF HCRT levels remain maximal if the wake period is extended by 4 h past the usual circadian phase of sleep-onset [259], and in rats, the magnitude of cFos induction in HCRT neurons in response to 2 h of sleep deprivation does not vary with circadian phase (e.g., Ref. [72]). Studies examining specific behavioral correlates of HCRT activity show that in cats, dogs, and hamsters, activity of HCRT cells reflects sensorimotor activation, rather than waking per se [145,239,246,254]. Consistent with these observations, SCN ablation in rats is associated with chronically low CSF HCRT-1 (comparable to the rest phase in intact rats; [49]), despite the lack of change in total daily wake time. Low HCRT-1 is likely due to the hypoactivity that is typical of SCN-ablated rats. CSF HCRT-1 levels are fully restored in SCN-ablated rats following 2-h sleep deprivation [50]. Although the ‘gentle handling’ procedure used to prevent sleep in that study does not require the rats to move continuously, it almost certainly increases general activity above baseline levels. A critical role for locomotion as a condition for detecting significant HCRT cellular activation may also explain why rats placed in a bright, noisy, novel open field during their usual sleep period exhibit arousal (e.g., stress responses) but no induction of cFos in HCRT cells [71]; bright light and noise can be expected to produce freezing behavior, and to inhibit rather than stimulate locomotor activity in a novel environment.

Three studies have examined the role of HCRT neurons in the expression of hyperactivity that normally occurs in small rodents during acute fasting or restricted daytime feeding. Mice with orexin/ataxin-3-3 genetic ablations of HCRT neurons fail to exhibit increased activity during acute food deprivation [257], and exhibit a significant reduction in the magnitude of the food anticipatory circadian activity rhythm that emerges when mice are fed once/day at a fixed time [4,153]. These mice also fail to show increased wake bout duration prior to mealtime, but this may be secondary to loss of a locomotor facilitating input normally provided by HCRT cells (e.g., wake bouts significantly increase when intact mice have free access to a running wheel [248]). Notably, circadian food anticipatory locomotor rhythms are not eliminated in the orexin/ataxin-3-3 mice, or in rats with HCRT2-sap lesions [158]. This places HCRT neurons on the output side of the circadian mechanism responsible for entrainment to feeding schedules, rather than on the input (entrainment) side or within the timekeeping loop.

These results, in aggregate, support a general hypothesis that HCRT neurons participate in promoting extended behavioral states and a level of arousal appropriate to
homeostatic and circadian factors regulating behavior. On the homeostatic side, a primary function of HCRT neurons may be to mediate metabolic influences on behavioral arousal and locomotion [152,159,230]. On the circadian side, HCRT neurons are essential for consolidating sleep and wake episodes without being required for the primary segregation of these episodes according to circadian phase. The effect of HCRT activation on consolidation of waking may be at least in part secondary to a role in facilitating motor output.

Specific monoamine cell groups downstream from HCRT neurons are also not essential for circadian timing of behavioral states. Despite evidence that histaminergic neurons in the TMN mediate the induction of arousal by intraventricular HCRT injections [97] and promote nocturnal wheel running [1], absence of histamine signaling in histidine decarboxylase knockout mice or destruction of histaminergic neurons in rats by HCRT-saporin neurotoxin has little effect on circadian sleep–wake or general activity rhythms [1,83,84,179]. Monoaminergic arousal cell groups in the brainstem (raphe, LC, and LDT/PPT) are also not necessary for circadian rhythms of sleep–wake; circadian rhythms of EEG-defined sleep and wake states recover in the forebrain of rats with high mesencephalic brainstem transections that disconnect these cell groups from HCRT inputs and forebrain targets [90]. Circadian organization of sleep–wake states is lost only when the posterior hypothalamus (likely including the TMN) is also ablated following brainstem transections, and in that preparation, the normal low-voltage EEG ‘wake’ state itself is absent [168]. Brainstem transections combined with large preoptic area lesions severely reduced the high-voltage EEG ‘sleep’ state, and circadian cycling, but these lesions may have also damaged the SCN [168].

5.6. SCN efferents to hypothalamic sleep–wake circuits: interim summary

The severity of lesion effects on circadian organization of sleep–wake can be tentatively rank ordered as follows: SCN > sPVZ > DMH, with more modest and less consistent results for preoptic area lesions, and minor or no effects with specific lesions further downstream, unless toxins (e.g., HCRT2-saporin) or lesions and transections are used sufficient to destroy large areas in combination. It may be that beyond the sPVZ, or perhaps the DMH, there is sufficient dispersal of circadian timing information to sustain significant circadian organization of behavioral state in the absence of any one component of the sleep and arousal systems. This conclusion remains tentative until the effects of complete lesions of these components, particularly those in the preoptic area, have been thoroughly characterized in both LD and DD.

Early neural systems models of sleep–wake regulation postulated interacting sleep and wake ‘centers’ in the anterior and posterior hypothalamus, respectively (e.g., [224]). This general concept has received considerable empirical validation. Reciprocally inhibitory pathways have now been demonstrated between the sleep-active cell groups in the POA, and wake-active, arousal promoting cell groups in the lateral and posterior hypothalamus and brainstem monoaminergic nuclei [79,149,203]. These pathways have been suggested to function analogously to an electronic ‘flip-flop’ circuit [203]. Activity on each side of the flip-flop is self-reinforcing, as this enhances inhibitory output to the opposing side, thereby reducing inhibitory feedback. In the context of sleep–wake regulation, this is conceptualized as a critical mechanism for ensuring stability of behavioral state, and for facilitating rapid switching among behavioral states at transition times when other factors (e.g., circadian and homeostatic) converge to tip the balance of activity from one to the other side of the circuit. The HCRT system appears to play a critical role in this mechanism [160]. Clearly, the circadian clock must also influence weighting within the flip-flop to produce the extended sleep–wake phases of the subjective day and night. The circadian influence may be strengthened by the presence of SCN outputs to both sides of the proposed circuit, either simultaneously pressing down on one side of the flip-flop while pushing up on the other, or sequentially pressing down or pushing up on one side, and then switching to the other side. This prediction could be tested by modeling, but ultimately electrophysiological and other empirical approaches will be needed to answer these questions; are SCN outputs to sleep–wake circuits silent during either the sleep or wake phase of the circadian cycle, or is there is continuous modulation, on one or both sides of the flip-flop circuit?

The neuroanatomical evidence for monosynaptic and polysynaptic pathways from the SCN to sleep-active and wake-active cell groups is consistent with the view that the circadian clock promotes both sleep and wake, at different phases of its daily cycle. It has been suggested that the role of the SCN in this circuit may be more heavily weighted toward stimulation of arousal, because DMH lesions in rats attenuate circadian sleep–wake rhythms and modestly increase total sleep time [35]. However, this argument is not compelling; if SCN output preferentially served to enhance arousal, then SCN lesions should also increase sleep time in rats, but they do not. The effects of DMH lesions on both circadian organization and total sleep time indicates only that the DMH is an important interface between the circadian clock and the wake-promoting side of the sleep–wake system.

5.7. SCN efferents regulate endocrine and autonomic correlates of sleep and wake

SCN outputs conferring circadian rhythmicity on effectors for specific endocrine and autonomic correlates of sleep–wake states have also been identified [12,26,27,112, 113,115,222]. SCN projections to the hypothalamic PVN/
DMH are of particular importance, as first demonstrated for circadian control of the pineal hormone melatonin [163]. The results of recent work on this polysynaptic pathway from SCN to pineal can be taken to illustrate two principles. One principle is that the SCN may drive at least some circadian endocrine rhythms by alternating between inhibitory and excitatory output at different circadian phases. Although the majority of SCN neurons are GABAergic [31,164], a significant proportion of peptidergic SCN neurons do not colocalize GABA [31], and may be glutamatergic [37,39,95]. In the case of melatonin secretion, experiments with GABA and glutamate antagonists suggest that the SCN, by outputs to the PVN, exerts a GABA-mediated inhibition of secretion during the subjective day, and a glutamate-mediated excitation of secretion during the subjective night [114,180,181]. Excitatory and inhibitory SCN outputs are also implicated in regulation of cortisol secretion and other autonomic and neuroendocrine functions (e.g., Refs. [38,39,111,198–200]).

A second principle suggested by these studies is that changes in the mean level of a regulated variable after SCN ablation may provide a misleading picture of SCN control of that variable. SCN ablation reduces melatonin secretion to ∼20% of the usual nocturnal level [180]. This supports the pharmacological evidence that the SCN has an excitatory effect on melatonin secretion during the subjective night. However, it belies other pharmacological evidence that the SCN inhibits melatonin secretion during the subjective day. Clearly, caution is warranted in drawing inferences from lesion results alone.

5.8. SCN diffusible factors actively inhibit arousal and locomotion in nocturnal rodents

Classical synaptic transmission is implicated in SCN control of sleep–wake and its neuroendocrine correlates (see also [46,151,210]), but there is now clear evidence that the SCN also regulates behavioral state by diffusible factors that can inhibit arousal and activity. The existence of inhibitory signals was initially inferred from analysis of activity patterns of Syrian hamsters bearing SCN transplants. SCN transplants can restore circadian rest–activity rhythms in hamsters with total SCN ablations (e.g., Refs. [135,187]). If the SCN ablation is partial, two rhythms may appear, with distinct free-running periods corresponding to the genotypes of the lesioned host and SCN donor animals (e.g., a wildtype host with an ~24-h circadian period and a ‘tau’ mutant donor with an ~20-h or 22-h circadian period; [242]). The two rhythms exhibit striking interference patterns, such that wheel running is expressed only when both circadian clocks signal subjective night. If either signals subjective day, wheel running is absent. Similar, albeit less dramatic examples of inhibition can be seen in the uncoupled activity components of diurnal tree shrews recorded in constant dark [149]. These results demonstrate that the expression of circadian rest–activity cycles reflects alternating excitatory and inhibitory effects of SCN output on locomotor activity [242].

The paracrine nature of this output was suggested by recovery of circadian rhythms in cases where the transplanted tissue did not appear to make axonal connections. More direct evidence was obtained by the use of semi-permeable polymer capsules that prevented axonal outgrowth from SCN grafts yet permitted restoration of rhythmicity [219], indicating that circadian clock-dependent excitation and inhibition of activity is mediated at least in part by diffusible molecules (see also [240]).

So far, two candidate molecules have been discovered. Transforming growth factor-α (TGF-α; [128]) and prokinecin-2 (PK-2; [33]) are both expressed by SCN cells during the subjective day and bind to receptors in major target areas of the SCN, including the sPVZ (TGF-α receptors) and DMH (PK-2 receptors). In Syrian hamsters, continuous intracerebroventricular infusion of TGF-α tonically suppresses wheel-running behavior and, like sPVZ lesions, eliminates the daily sleep–wake rhythm without altering sleep duration [128]. Single microinjections of TGF-α or PK2 acutely suppress locomotor activity during the subjective night in hamsters [128] and rats [33], respectively. Surprisingly, TGF-α has been localized primarily to glial fibrillary acidic protein-positive cells within the SCN, suggesting a role for glial cells as a source of diffusible SCN outputs [136]. This possibility is given credence by evidence that pure cultures of astrocytes can express circadian rhythms of Per1 or Per2 activity [184].

These results suggest that circadian clock-dependent suppression of activity and arousal during the daily rest period in nocturnal rodents is mediated by TGF-α and PK2 acting on receptors in SCN target areas (sPVZ and DMH) already identified as important for circadian regulation of sleep–wake. Whether these factors mediate restoration of circadian function by SCN transplants remains to be determined. To date, diffusible factors that promote arousal and activity have not been found.

5.9. Neural basis of diurnality: preliminary observations

Diurnal mammals have been under-represented in neurobiological analysis of the circadian and sleep–wake systems, and the neural basis for different circadian chronotypes still remains to be elucidated (for review, see Ref. [221]). The circadian rhythms of neural and metabolic activity and clock gene expression in SCN neurons are similarly phased relative to the LD cycle in nocturnal and diurnal mammals [28,44,86,104,116,126,130,201,209]; therefore, inverse phasing of sleep–wake and other rhythms is probably not due to inverse phasing of clock cells (but note that one study of the diurnal hystricomorph rodent, Octodon degus, found no circadian variation of SCN single unit activity [106], and another study has reported different phasing of circadian rhythms of VIP and GRP mRNA expression in the SCN of the nocturnal mouse Mus musculus and the diurnal rat.
Arvicanthus ansorgei [45]). Assuming that the phase of SCN clock cells relative to local time is the same in animals of different chronotypes, then differences in the timing of overt rhythms must be controlled downstream from the core of the clock, either in the composition of SCN output signals (e.g., different ratios of inhibitory and excitatory SCN transmitters), or in one or more targets of SCN efferents (e.g., presence of interneurons that invert SCN signals in one chronotype relative to the other). A candidate site for circadian signal inversion is the region of the sPVZ immediately dorsal to the SCN [221]. Neurons here exhibit a circadian rhythm of cFos expression that is differentially phased in the diurnal unstriped Nile grass rat, Arvicanthus niloticus, relative to Norway rats ([175]), and this rhythm persists in DD only in the grass rat [211]. The sPVZ projects to sleep–wake circuits in parallel with the SCN, and may be responsible for modifying SCN timing signals to produce nocturnal and diurnal phenotypes.

Despite the marked difference in the timing of sleep–wake rhythms in nocturnal and diurnal animals, there is no reason, a priori, to assume that the underlying mechanism involves a different pattern of connections between the SCN and sleep–wake circuits. Indeed, there are some species that switch between nocturnality and diurnality with time of year or other factors (e.g., Refs. [41,177,221]), and, whether or not this reflects a true change in pacemaker phase, as opposed to a direct behavioral response to one or more environmental stimuli, this is not likely to be associated with spatial re-organization of SCN projections. At present, a comprehensive mapping of SCN projections is not yet available for any of the diurnal rodent models currently being used in sleep and chronobiology research. There is evidence from studies of diurnal grass rats and degus that is consistent with a role for VLPO and HCRT neurons in sleep and wake functions, respectively [80,143,172–174]. There is preliminary evidence for PK-2 fibers and receptors in the hypothalamus, including the SCN, of grass rats, with a regional distribution similar to that in nocturnal mice [140]. Analysis of behavioral effects of PK-2 in the grass rat are awaited with interest; does this molecule signal

Effects of PK-2 in the grass rat are awaited with interest; does this molecule signal that circadian rhythms of neuroendocrine variables are driven by alternating excitatory and inhibitory outputs from the SCN [27]. The available evidence also indicates that both classical synaptic transmission and paracrine signaling participate in generating the circadian sleep–wake cycle, and that the rest portion of the cycle reflects, at least in part, active SCN-dependent suppression of activity and arousal. Future efforts will be needed to further characterize the role of these and other potential hypothalamic and extra-hypothalamic neural pathways and diffusible signals from SCN to sleep–wake circuits. Lesions of sleep- or wake-promoting cell groups may attenuate but usually do not eliminate the circadian sleep–wake cycle, suggesting multiple pathways of circadian control. These pathways may be grossly redundant, or may make specialized contributions to the daily circadian waveform as part of a network of interacting components (e.g., the hypothesis of HCRT-mediated circadian alerting late in the wake phase). The full array of functional and anatomical experimental approaches will be needed to determine how differently phased SCN output cells are mutually coordinated, whether the post-synaptic effects of these outputs vary with circadian phase, how the circadian signal is integrated with other influences (e.g., homeostatic, environmental) on sleep–wake executive circuits, and where among these mechanisms the nocturnal and diurnal sleep–wake phenotypes emerge.

6. Hypersomnia in SCN-ablated squirrel monkeys, revisited

Unlike in nocturnal rats, SCN ablation in diurnal squirrel monkeys produces an apparent hypersomnia, characterized by markedly increased total daily sleep time, fragmentation of sleep and wake bouts [69], and short sleep latencies when the lights are turned off (see Fig. 4 in Ref. [61]). From these observations, it has been inferred that the SCN actively stimulates arousal but does not actively promote sleep, and that hypersomnia is caused by loss of this circadian alerting signal. Is this conclusion logically necessary, or might there be other causes for hypersomnia? If there are other causes, can these account for the apparent species differences in the effects of SCN ablation on total daily sleep time?

Sleep expression is controlled by four primary factors; the circadian clock, intrinsic sleep ‘need’ (reflected in the homeostatic properties of sleep and determined by recent sleep–wake history), other needs or ‘drives’ (e.g., hunger, thirst, reproduction, migration) and environmental conditions (e.g., light, sound, temperature, social stimuli). If we assume for the sake of parsimony (or debate) that circadian regulation of sleep–wake states is continuous or bidirectional in rats, humans, and monkeys alike, that leaves three factors to examine as possible causes for the differential effects of SCN ablation on total sleep time in rats and monkeys.

Competing physiological needs do not seem to be important, because SCN ablation does not affect food
intake, body weight, fluid balance, or mean body temperature in rats or monkeys. Sleep ‘need’, by comparison, is more difficult to evaluate. SCN-ablated monkeys fall asleep rapidly in the dark, and sleep several hours longer each day than do intact monkeys. This could be because a portion (i.e., the SCN) of the arousal system is missing. An alternative conception is that hypersomnia reflects an increased need for sleep time. If so, why? One possibility, examined in section 3.2 with respect to reduced SWA in SCN-ablated rats, is that SCN ablations reduce the restorative value of sleep, thus increasing the amount of sleep required to achieve functional recovery. The restorative value of sleep has been linked to both sleep intensity (e.g., as operationalized by SWA and arousal threshold) and sleep continuity [20]. In both respects, SCN-ablated squirrel monkeys appear disadvantaged. Although SCN-ablated monkeys sleep nearly 4 h more each day, the total daily amount of high intensity, slow-wave-enriched NREMS (so-called ‘SWS2’, defined as NREMS in which 25% or more of a 30-s scoring epoch displays EEG waves of $< 3$ Hz and $> 75\,\mu$V; [69]) remains constant, i.e., it is significantly reduced as a percentage of total sleep time; thus, average sleep intensity, by definition, is reduced. Notably, the arousal threshold is apparently also reduced in SCN-ablated monkeys, consistent with a lower average sleep intensity (unpublished data cited in Ref. [69]). Sleep extension and sleep restriction studies in humans indicate that total daily SWA is homeostatically conserved [23,56,74]. It may be that reduced sleep intensity merely reflects sleep satiation due to the increase in daily sleep amounts. However, it cannot be ruled out that the arrow of causality points the other way; a primary deficit in sleep intensity may increase total sleep time, to meet homeostatic requirements. Consistent with this idea, humans selectively deprived of slow-wave-enriched stages 3–4 during daytime sleep after a night awake exhibit a significant increase in sleep duration, i.e., sleep duration expands when its intensity (operationally defined by SWA) is experimentally attenuated [85]. In another study, the duration of nocturnal sleep was not similarly increased by 3–5 h of selective stage 3–4 deprivation initiated at sleep onset, but the lost SWA in those subjects was rapidly and entirely recouped during the recovery portion of the night, and this would have eliminated any need for additional sleep [56]. The restorative value of sleep in humans, as reflected in daytime sleepiness, is also reduced by experimental sleep fragmentation (e.g., repeated nocturnal awakening; [20]). Apparent hypersomnia in SCN-ablated monkeys may thus also be related to their fragmented sleep; the average sleep bout length in SCN-ablated squirrel monkeys is approximately half of that exhibited by intact monkeys in their usual sleep phase [69].

Changes in the sleep of squirrel monkeys following SCN ablation have been interpreted as evidence that the sole function of the SCN is to promote wakefulness but, logically, the results are also compatible with a role for the SCN in promoting sleep. One important function of the SCN may be to facilitate sleep recovery processes by promoting sleep intensity and sleep continuity. This alternative conception could be tested by subjecting intact squirrel monkeys to partial SWS2 deprivation, with sufficient awakenings to mimic the fragmented sleep patterns of SCN-ablated monkeys. If reduced sleep intensity and continuity accounts for a significant portion of hypersomnia in SCN-ablated monkeys, then intact monkeys may also exhibit an increase in total sleep time. Due to the presence of the SCN in these animals, sleep expression would presumably be modulated by circadian phase (e.g., Ref. [122]), and the total increment in daily sleep would no doubt be less severe.

This hypothesis of hypersomnia due to reduced sleep intensity does not explain the differential effects of SCN ablation on total daily sleep time in rats and monkeys, but then, neither does the Opponent Process account. Sleep recovery processes in rats may be less dependent on sleep continuity, because the rat is already adapted to a more polyphasic sleep process, due to competing metabolic needs (a high metabolism requiring more frequent feeding, drinking, and elimination) or other factors. These same factors may enforce a ceiling on total sleep time, that, on an hourly basis, is capped lower in the absence of circadian organization of physiology than it is during the rest phase of the circadian cycle in intact rats. It may be that we will not fully understand the effects of SCN ablation on daily sleep amounts until we have a better grasp of the physiological functions of sleep that underlie its homeostatic properties, and the relation of these functions to presumed correlates of sleep intensity (for discussion, see Ref. [189]).

The remaining factor to consider in this re-examination of total daily sleep time in SCN-ablated squirrel monkeys is the environment. Environmental stimuli can have direct effects on behavioral state. In nocturnal rodents, light inhibits activity and promotes sleep, while dark promotes activity and waking [21,190]. These direct effects are as strong or stronger when the SCN are removed, as demonstrated using ultradian LD cycles (e.g., 2 h light alternating with 2 h dark [6,191,220]). In diurnal animals, the opposite is true; light promotes arousal and dark promotes sleep (e.g., [79]). Under a 2 h:2 h LD cycle, SCN-ablated squirrel monkeys fall asleep rapidly at lights off, with a latency comparable to that in intact monkeys during their usual sleep period [61]. This has been interpreted as hypersomnia, and cited as a crucial piece of the evidence that the SCN normally promotes only arousal. However, what has been neglected in this interpretation is the behavioral response to the 2-h dark portion of this LD cycle. Apparently, SCN-ablated squirrel monkeys, by comparison with intact monkeys, also show a short wake onset latency in response to light (unpublished observations cited in Ref. [61]). Considered in isolation, this latter result could be taken to imply that SCN-ablated monkeys are less,
not more, sleepy than intact monkeys, and that the SCN normally promotes only sleep. Taken together, however, the results suggest that SCN ablation increases the sensitivity of the squirrel monkey to changes in environmental lighting, and thus do not constitute unambiguous support for a unidirectional model of sleep regulation.

Another aspect of the environment is its stimulus complexity. The sleep of SCN-ablated monkeys and rats has so far only been studied under controlled laboratory environments that lack opportunities for normal physical, sensory, and social stimulation. Confinement to an impoverished environment should, at least in the short term, favor sleep over wake. In humans, exposure to a 14-h night (lights-off) increases the duration of the daily sleep phase by ~3.5 h and stably increases total sleep by ~1 h ([247]; see also Ref. [62]). During the mid-day, boredom exacerbates sleepiness and reduces reaction times [146]. In a 60-h forced bed rest protocol in which reading, writing, television, music, and exercise were prohibited and social contact restricted to meal delivery (i.e., a very low stimulation environment), subjects slept an impressive 47.6% of the time (i.e., nearly identical to SCN-ablated squirrel monkeys; [29]). There is no denying that sleep in humans is substantially self-limiting; we do not sleep continuously during forced bed rest or solitary confinement. Nonetheless, sleep is promoted and can significantly expand under monotonous conditions, even with the circadian clock intact. Squirrel monkeys in sleep and circadian rhythm studies lived alone in small enclosures with little to do during recording sessions. The lack of both environmental and circadian clock stimulation of arousal in SCN-ablated monkeys may cause sleep time to rise despite the loss of an SCN-dependent circadian sleep-promoting factor. There is no reason to expect a balanced effect on sleep–wake amounts if circadian sleep- and wake-promoting factors are both absent and the monkeys are tested in environments lacking environmental wake-promoting factors. Sleep inertia (the low arousal state immediately following sleep; [190]), unchecked by environmental stimulation, would be expected to further favor sleep over wake.

Note that this is an interaction hypothesis; intact and SCN-ablated monkeys are hypothesized to be differentially affected by confinement to a monotonous environment because of a non-additive interaction between endogenous circadian and environmental alerting factors. According to this hypothesis, sleep in intact and SCN-ablated monkeys might be substantially more similar in a socially enriched environment. To explain the difference between SCN-ablated rats and squirrel monkeys, it is necessary to consider the possibility that squirrel monkeys are more sensitive than are rats to environments lacking normal opportunities for physical, sensory, cognitive, and social stimulation. This is not out of the realm of possibility.

A variation on this interaction hypothesis is that hypersomnia in SCN-ablated squirrel monkeys reflects not just boredom, but psychopathology. While working with squirrel monkeys in temporal isolation studies, it was the intuition of this observer that these highly social primates might become psychologically depressed by solitary confinement. In humans, endogenous depression is typically associated with reduced SWA, particularly in the first sleep cycle of the night, and depression, while often associated with insomnia, can also be characterized by hypersomnia (reviewed in Ref. [15]). Some squirrel monkeys clearly do not tolerate solitary confinement, become hypoactive, and have to be returned to the group colony (Mistlberger, unpublished observations, 1986–88). Most squirrel monkeys can be maintained in isolation, but sleep studies reveal a delayed peak of EEG SWA during nocturnal sleep [123] that is reminiscent of the sleep EEG in human depression [15]. To use psychopathology as an explanation for hypersomnia in SCN-ablated monkeys, an interaction effect must again be invoked. Is it possible that the stress of surgery and experimentally-induced brain damage in squirrel monkeys sets the conditions for adverse physiological responses to social isolation?

The role of the environment and of changes in the restorative value of sleep as causal factors underlying the apparent hypersomnia of SCN-ablated squirrel monkeys is speculative, but testable. Regardless of the ultimate merits of these alternative hypotheses, it is clear that this apparent hypersomnia is ambiguous in what it reveals about SCN control of sleep–wake, and does not logically preclude a role for the SCN in actively promoting sleep.

7. Anatomical specificity of SCN ablation effects

Is it possible that the effects of SCN ablations on total daily sleep time are caused by incidental damage to other hypothalamic nuclei in the SCN vicinity, or to retinal pathways conveying photic input to these nuclei? This is considered unlikely. SCN ablations should cut some retinohypothalamic fibers that innervate other parts of the hypothalamus. However, as already noted, SCN ablation does not eliminate photic masking of sleep–wake states [192,220], indicating that the retinal pathways and hypothalamic and extra-hypothalamic areas mediating the behavioral response to light and dark are substantially intact. Also, there is no evidence for a relation between SCN lesion size and total daily sleep time. In fact, typical SCN lesions produce some damage to the ventral sPVZ but are rarely so large as to damage either the VLPO or the DMH. Complete SCN lesions can be quite discrete, producing minimal damage to non-SCN areas (e.g., Refs. [69,255]). It is particularly doubtful that the SCN lesions made in the squirrel monkey study [69] could have damaged wake promoting cell groups in the basal forebrain, DMH, or posterior hypothalamus that might explain hypersomnia. The only caveat to this conclusion is that functional
mapping of sleep–wake neural circuits has not been done for squirrel monkeys, and control lesions were not done to rule out the possibility that a wake promoting area resides close to the SCN. In the case of rats, control lesions in the peri-SCN region have been done; as noted, lesions confined to the sPVZ, like total SCN lesions, do not affect total sleep time [138].

8. Opponent processes: staying clear on the concept

Studies of nocturnal Norway rats indicate that the SCN in this species actively promotes both sleep and wake. The overall lack of effect of SCN lesions on total daily sleep time in this species further suggests that these sleep and wake promoting actions must be approximately equivalent. There is no reason to assume that the same is true of all species; the SCN could exert a stronger positive drive on wake in some species, particularly those with extended, consolidated daily wake periods, in which case SCN ablation may result in increased daily sleep time. However, if by other behavioral or neurobiological analyses the SCN are shown to also have an active sleep promoting role in these species, then the lesion data cannot be construed as support for the Opponent Process model (Model 1). As formulated, the Opponent Process model only has meaning if the role of the clock is restricted to an alerting function that opposes a sleep-generating factor. If the SCN also has an active role in promoting sleep, then the data support Model 3 (continuous or bidirectional control of sleep–wake by the clock), not Model 1 (discontinuous or unidirectional control).

9. Conclusions

The advent of modern genomics heralds an exciting new phase in discovery research on the regulation and functions of sleep and circadian rhythms [125,141,213]. Mutations and knockouts of putative circadian clock genes have already been identified that increase (arrhythmic Cry1,2\(^{-/-}\) knockout mouse; [253]), decrease (the Clock mutation; [170]), or have no effect on total daily sleep time (arrhythmic Per1,2,3\(^{-/-}\) knockout mouse; [216]). Given the apparent bidirectional control of sleep and wake by the circadian clock in rodents, a diversity of effects might well be anticipated, if different clock genes or clock-controlled genes participate in different output functions of the clock. The points made here concerning the interpretation of sleep changes in SCN-ablated animals are also relevant to sleep in animals bearing genetic manipulations. To interpret changes in sleep time, it will be important to probe sleep–wake regulation using selective and total sleep deprivation, and to consider environmental and behavioral variables that may influence the expression of sleep independently of homeostatic or circadian factors.

The idea that sleep is regulated by opponent processes is attractive in its simplicity, and this no doubt accounts for the dissemination of this concept in scientific, didactic, and popular writing on sleep. However, a case can be made that the data on which this model relies have been overinterpreted. Moreover, this interpretation implies significant species differences in the mechanisms by which the circadian clock exerts control over behavioral state. Given the small number of species that have been studied in any detail, species differences cannot yet be ruled out. But, if there are differences, these may ultimately prove to be minor variations in the balance between active wake and sleep promoting effects of the clock, as opposed to major qualitative differences in the organization and nature of SCN outputs. The data and logical arguments considered here support a conclusion that circadian clock-dependent alerting is only half the story of sleep regulation in mammals, and lead to a prediction that circadian clock outputs will be found to actively engage the sleep–wake system during both the sleep and wake portions of the circadian cycle.

Acknowledgments

Supported by NSERC, Canada. I thank Drs. M.C. Antle, T. deBoer, C.I. Eastman, J.D. Miller, N.F. Ruby, T. Scammell, K. Semba, and D.K Welsh for comments and discussion.

References


J.P. Nixon, L. Smale, Individual differences in wheel-running rhythms are related to temporal and spatial patterns of activation.


