Neurobiology of Sleep

Brandon S. Lu, MD, MSa, *, Phyllis C. Zee, MD, PhDb

Sleep is an evolutionarily conserved process that occupies approximately one-third of a human’s life. Although the exact functions of sleep remain elusive, studies of sleep deprivation have shown impairments in cognitive and physical performance, and chronic short-sleep duration has been associated with numerous cardiometabolic disturbances, including hypertension, diabetes, and even mortality, emphasizing the importance of sleep for health and performance. Thus, knowledge of the neurobiology of sleep and wake regulation is essential to medicine.

Our understanding of sleep mechanisms has improved drastically in recent years, largely aided by improved methods in molecular and cellular experimental techniques that have allowed direct access to sleep-wake centers in the brain. For example, in vivo microdialysis was used to measure extracellular adenosine concentrations in the basal forebrain of free, living animals to better delineate adenosine’s role in sleep. Similarly, lesion of ventrolateral preoptic nucleus (VLPO) cells with direct microinjection of an acid identified distinct cell groups with primary effects on the regulation of rapid eye movement (REM) sleep and non-REM (NREM) sleep.

This article focuses on the how of sleep, not the why of sleep. The authors discuss the wake and sleep centers, the transition between wake and sleep, regulation of REM and NREM sleep, and the homeostatic and circadian regulations of sleep and wakefulness.

WAKE-PROMOTING SYSTEMS

After the initial recording of brain electrical activity by Hans Berger in the 1920s, a major advancement in sleep neurobiology was reached when von Economo described in detail the symptoms and pathology of encephalitis lethargica. He hypothesized that the posterior hypothalamus and rostral midbrain contained centers of wakefulness (lesions in these areas led to excessive sleepiness) and the anterior hypothalamus controlled sleep (lesions led to prolonged insomnia). Moruzzi and Magoun further defined the sleep-wake transition when they demonstrated that stimulation of brainstem reticular formation evoked a generalized desynchronization of electroencephalogram (EEG) activity, simulating arousals from sleep. They hypothesized that a series of reticular relays distributed through the center of the brainstem projects to the basal forebrain and participates in the regulation of wakefulness.

Much work has been performed in the last half decade to identify several distinct systems involved in the control of the wake state.

Cholinergic Systems

The “ascending reticular activating system” initially described by Moruzzi and Magoun contains two main branches. The first branch consists of cholinergic neurons originating from the laterodorsal tegmental (LDT) and pedunculopontine nuclei (PPT) of the dorsal midbrain and pons. Dorsal projections from these centers densely innervate the thalamic relay nuclei and...
the thalamic reticular nuclei, with neurons displaying high firing rates during wakefulness and REM sleep.11,12 From the thalamus, arousal signal is carried by the thalamocortical tract to activate the cortex, resulting in the desynchronized, low-amplitude EEG signals seen during wakefulness and REM sleep.13,14

A ventral projection from the LDT and PPT carries cholinergic neurons to the basal forebrain, where areas such as the substantia innominata and the medial septum and diagonal band of Broca relay the signals to the cerebral cortex.15 The basal forebrain has high activity during wake and REM sleep, contributing to EEG desynchronization particularly gamma activity (30–60 Hz).16 Lesions to animal basal forebrain area produced slow delta waves recorded from the cortex during activity and immobility, further suggesting the importance of this area in cortical activation.17 Aside from its projections to the cerebral cortex, the basal forebrain is also a source of cholinergic activation to the hippocampus and amygdala.18

**Monoaminergic Pathways**

The second branch of the reticular activating system carries mostly monoaminergic neurons to the lateral hypothalamus (LH) and basal forebrain, and ultimately to the cerebral cortex.15 Monoaminergic nuclei located in the upper brainstem and caudal hypothalamus, which send fibers rostrally, include the locus coeruleus (LC), median and dorsal raphe (DR), tuberomammillary (TMN), substantia nigra (SN), and ventral tegmental area (VTA).

**Norepinephrine**

The LC is the main source of brain norepinephrine and carries projections to subcortical relay stations (ie, thalamus and hypothalamus) as well as the cortex. In contrast to the cholinergic system, LC neurons discharge highly during wake, less during NREM sleep, and are virtually off during REM sleep.20,21 Increasing LC neuronal activity in experimental studies promoted wakefulness, whereas LC inactivation led to a shift of EEG from low-amplitude, high-frequency to large-amplitude, slow-wave activity, confirming the role of LC in wakefulness.22,23

The LC-norepinephrine system also functions in the regulation of behavioral state. It displays pronounced sensitivity to stressors and has been implicated in stress-related arousal. Consistent with this suggestion, bilateral suppression of LC discharge blocked stressor-induced EEG activation in halothane-anesthetized rats.24,25

**Serotonin**

Serotonergic—5-hydroxytryptamine (5-HT)—neurons originate in the DR nuclei and, similar to the norepinephrine neurons of the LC, are very active during wake, less active during NREM sleep, and silent during REM sleep.26 Our understanding of the sleep-wake property of 5-HT has changed over time. Earlier works pointed to 5-HT as a somnogenic agent based on studies which showed that 5-HT increased sleep and inhibition of 5-HT synthesis suppressed NREM and REM sleep.27 Furthermore, perfusion of a 5-HT1A agonist into the DR-increased REM sleep—a finding consistent with likely activation of presynaptic autoreceptors leading to decreased 5-HT neurotransmission.28 More recently, however, several studies have shown that 5-HT promotes wakefulness by using various 5-HT receptor agonists/antagonists, 5-HT precursors, or reuptake inhibitors to increase quiet waking and decrease REM sleep.29–31 Like norepinephrine, 5-HT also regulates behavioral state.

**Histamine**

The TMN in the posterior hypothalamus is the sole source of histamine (H) in the brain and it sends projections to the entire central nervous system. Active mainly during the waking period, the TMN has little if any activity during sleep, particularly in REM sleep.32,33 In studies of animals and from clinical experience, drugs that block histamine H1 receptors (eg, diphenhydramine) increase both NREM and REM sleep.34 Different types of histamine receptor exist, however, providing distinct pharmacologic targets in the same system—H1 and H2 receptors are mostly excitatory in nature, whereas H3 receptors act as inhibitory auto- and heteroreceptors. In fact, H3 receptor antagonists are being developed for the management of narcolepsy.35

**Dopamine**

Dopamine (DA)-containing neurons that are relevant to the sleep-wake cycle are predominantly found in the SN and VTA. These nuclei have interconnections with many nuclei in the brainstem as well as the LH, basal forebrain, and the thalamus. Two subtypes of DA receptors have been cloned—the D1-like and D2-like subfamilies—with the D1 receptor being postsynaptic and its stimulation leading to behavioral arousal, increased wake, and decreased slow-wave sleep (SWS) and REM sleep. The D2 receptors are autoreceptors and postsynaptic receptors that have a biphasic response to agonists—low doses leading to decreased wake and increased SWS and REM sleep, and high doses having the...
opposite effect. Distinct from other monoaminergic neurons in the brainstem, DA-containing neurons do not change firing rate across sleep-wake state; rather, the burst activity of DA-containing neurons occurs in a temporal pattern, associated with reward, locomotion, and cognition. DA neuronal firing depends upon inputs from the prefrontal cortex, the LH, and cholinergic as well as other monoaminergic nuclei in the brainstem.

The wake-promoting effect of DA is readily apparent when one considers that amphetamines promote wakefulness by enhancing DA release and preventing its reuptake by DA transporter. Medications that block DA receptors, such as antipsychotics, often lead to sleepiness, and patients with Parkinson’s disease often exhibit excessive daytime sleepiness, which may be partially attributed to DA deficiency. Interestingly, dopamine agonists have been associated with sudden sleep attacks. The mechanism of this phenomenon is possibly due to down-regulation of DA input by the medications binding to presynaptic DA receptors. Taken together, these data clearly support the wake-promoting action of DA.

**Orexin/Hypocretin Pathway**

In 1998, two groups of scientists independently discovered orexin/hypocretin, a neuropeptide synthesized predominantly in the posterior and LH. Orexinergic neurons play an essential role for the stabilization of wakefulness by projecting and binding to its two receptors, OX1 and OX2, found diffusely throughout the central nervous system, including the cerebral cortex, forebrain, thalamus, and brainstem arousal nuclei. In addition to cortical activation, orexin stimulates somatic motor neurons and the sympathetic nervous system to maintain a waking state. It is thus not unexpected to find that orexin neurons discharge exclusively during wake and are off during both SWS and paradoxical sleep.

The clinical importance of orexin emerged when subsequent studies disrupting the orexin/hypocretin pathway using orexin knockout and hypocretin (orexin) receptor 2 gene (Hcrtr2 or OX2) mutation models produced animals with narcolepsy with cataplexy. About 90% of human narcoleptics with cataplexy are found to have low to undetectable levels of orexin in their cerebrospinal fluid, and human narcoleptics have a 90% reduction in brain orexin neurons, with evidence pointing to a degenerative process as the cause.

Of all the arousal systems discussed, the orexin/hypocretin system seems to have the most potent effect on maintaining wakefulness. Even after massive destruction of animal norepinephrine neurons in the LC, neither a comatose state nor a reduction of waking measured by cortical activation was seen. The same can be said for the cholinergic and other monoaminergic brainstem systems—suggesting a redundancy to those arousal systems. The consequence of a defective orexin system is well illustrated in narcolepsy. The orexin/hypocretin system thus performs its critical role of preventing sleep by its excitatory actions on other arousal systems and also providing excitatory input on cortical, motor, and sympathetic systems (Table 1).

**NREM SLEEP SYSTEMS**

The notion that sleep simply results from inactivity of the ascending arousal systems has proven to be erroneous with the discovery of specific brain regions that actively control NREM and REM

<table>
<thead>
<tr>
<th>Table 1 Activity of wake and sleep centers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wake</strong></td>
</tr>
<tr>
<td>LDT/PPT (Acetylcholine)</td>
</tr>
<tr>
<td>LC (Norepinephrine)</td>
</tr>
<tr>
<td>DR (5-HT)</td>
</tr>
<tr>
<td>TMN (Histamine)</td>
</tr>
<tr>
<td>SN/VTA (Dopamine)</td>
</tr>
<tr>
<td>Lateral Hypothalamus (Orexin/Hypocretin)</td>
</tr>
<tr>
<td>VLPO - cluster (Galanin &amp; GABA)</td>
</tr>
<tr>
<td>VLPO - extended (Galanin &amp; GABA)</td>
</tr>
</tbody>
</table>

Abbreviations: GABA, γ-aminobutyric acid; LDT, laterodorsal tegmental; PPT, pedunculopontine tegmental.

a high activity.
b low or negligible activity.
c activity.
The centers that facilitate sleep include the VLPO and the median preoptic nucleus (MnPN). These areas are active at the transition from waking to sleep and inhibit the firing of arousal centers for sleep initiation and maintenance.

**VLPO**

The VLPO is located in the anterior hypothalamus and corresponds to the lesioned area in von Economo’s patients with encephalitis lethargica who developed profound insomnia. Scientists have discovered that projections from the VLPO reach monoaminergic arousal centers and are active during sleep, using inhibitory neurotransmitters galanin and γ-aminobutyric acid (GABA). These cells form a dense cluster that has heavy innervations to the histaminergic neurons of the TMN, as well as a more diffuse extended part of the nucleus that provides output to the cholinergic LDT/PPT and monoaminergic LC and DR. To further define the function of the VLPO cluster, targeted lesions of the area led to a dramatic decrease in NREM sleep.7 The extended VLPO (eVLPO), on the other hand, showed activity that correlated with REM sleep (see **Table 1**). The orexin neurons of the LH also receive inhibitory input from the VLPO.

The activity of the VLPO is partially controlled by the same monoaminergic centers that it inhibits. All major monoaminergic nuclei send inhibitory projections to the VLPO so that the sleep and wake systems are reciprocally connected. There are mutual projections between the VLPO and the orexin neurons in the LH. However, the VLPO does not contain orexin receptors such that orexin neurons likely inhibit the VLPO via an indirect mechanism.

**MnPN**

Similar to the VLPO, the MnPN uses GABA to send inhibitory projections to many of the same arousal targets. The nuclei differ in the temporal pattern of discharge during routine sleep and following sleep deprivation. MnPN neurons tend to display highest activity early in NREM sleep with gradual decline through the sleep period, whereas VLPO neurons demonstrate sustained discharge. In experiments with sleep-deprived rats, it was found that MnPN showed peak activity during sleep deprivation, before the onset of recovery sleep. VLPO, in contrast, showed increased activity only after sleep onset. These data suggest a homeostatic regulatory role for the MnPN.

**THE SLEEP-WAKE TRANSITION (FLIP-FLOP SWITCH)**

The sleep and wake systems are mutually inhibitory and activity of one system will inhibit activity of the opposing system. Such a circuit is termed a flip-flop switch by engineers who designed such systems to be in one state or another, but never in-between (much like the on-off light switch). Because each state will reinforce its own stability when active (eg, the arousal system will inhibit the sleep system to decrease its own inhibitory feedback, and vice versa), the flip-flop switch is inherently stable and tends to avoid intermediate states (**Fig. 1**).

However, the flip-flop switch may be weakened and state transitions can occur frequently when one side becomes less able to inhibit the other side. An example is the suggested loss of VLPO neurons in the elderly that can lead to sleep

---

**Fig. 1.** Regulation of wake and sleep: flip/flop switch. The mutual inhibitory wake and sleep systems each reinforces its own stability when active. The orexin system is a stabilizer by its excitatory action on monoaminergic systems and inhibitory action on the VLPO. Solid arrows represent inhibitory action; open arrows; excitatory action. Thickness of arrows represent strength of relationship.
fragmentation at night and sleepiness or napping during the day, both commonly seen with aging. Orexin has been proposed as a stabilizer of the switch by virtue of its indirect inhibitory action on the VLPO. The orexin neurons actively reinforce the monoaminergic arousal tone by its excitatory activity, thus preventing the rapid and frequent transitions in behavioral states seen in narcoleptics who lack orexin.

**REM SLEEP SYSTEMS**

REM sleep was previously thought to be controlled by the interplay between the active pontine cholinergic nuclei and inactive monoaminergic nuclei that occurs during REM sleep. Indeed, this concept was supported by evidence that cholinergic stimulation promoted REM sleep and monoamine re-uptake inhibition (eg, antidepressants) reduced REM sleep. However, limited change in REM sleep was observed when either set of nuclei was selectively lesioned. More recent works have identified distinct REM-on and REM-off regions that also function in a flip-flop switch manner to regulate REM sleep (Fig. 2).

**REM-off Region**

By tracing the projections of the orexin neurons (excitatory and REM-inactive) and eVLPO (inhibitory and REM-active), scientists have identified a REM-off region as a crescent-shaped arc of tissue in the mesopontine tegmentum, consisting of the ventrolateral part of the periaqueductal gray matter (vlPAG) and the lateral pontine tegmentum (LPT). Consistent with an earlier report, lesions at these sites led to increased amounts of REM sleep during both light and dark periods in rats. In addition, bouts of cataplexy-like state with desynchronized EEG and atonia were also seen.

**REM-on Region**

The efferent projections of the REM-off region and c-Fos protein expression (an immediate early gene that is upregulated in response to many extracellular signals) were examined during enhanced-REM sleep to determine the site of a REM-on region. It was seen that the REM-off region provided heavy projection to the sublaterodorsal nucleus (SLD; also known as the subcoeruleus area or peri-locus coeruleus alpha) and the periventricular gray matter. This finding was consistent with earlier studies which showed that neurons in the SLD are REM-active and stimulation of the region increased REM-sleep–like behavior.

The ventral SLD (vSLD), an area below the periventricular gray matter, has been shown to be responsible for the atonia seen during REM sleep. In fact, animals with lesion of the vSLD were noted to exhibit complex motor behaviors during REM sleep. More extensive studies have demonstrated that the vSLD produces atonia by means of direct glutaminergic spinal projections to interneurons that inhibit spinal motor neurons by glycine and GABA neurotransmitters.

**Regulation of REM Sleep**

Akin to the transition between wake and sleep, the control of REM sleep is regulated by a system that resembles a flip-flop switch. The REM-on and REM-off neurons of the SLD/subcoeruleus region and vlPAG-LPT, respectively, mutually send reciprocal GABAergic inhibitory neurons to the other. The cholinergic LDT/PPT have been proposed to excite the REM-on region and inhibit the REM-off region, while the orexin and monoaminergic systems may both excite the REM-off region. Solid arrows represent inhibitory action; dashed solid arrows, inhibitory (hypothesized) action; dashed open arrows, excitatory (hypothesized) action. Thickness of arrows represents strength of relationship. LPT, lateral pontine; SLD, sublaterodorsal.
reciprocal GABAergic inhibitory neurons to the other. As each side inhibits the other, it also disinhibits and reinforces its own firing, stabilizing the switch.

To integrate the roles of the pontine cholinergic and monoaminergic nuclei (previously thought to be the REM switch) with the recent concept of the REM flip-flop switch, it might be more appropriate to think of the pontine cholinergic and monoaminergic nuclei as REM modulators. For example, the REM-active cholinergic neurons of the LDT/PPT may inhibit the LPT (cholinergic agonists injected into the LPT were shown to cause a REM state) and also directly excite the SLD/subcoeruleus REM-on region. Similarly, 5-HT (DR) and norepinephrine (LC) neurons may actively excite the vIPAG-LPT REM-off neurons to prevent sudden, unexpected transitions into REM sleep. To reinforce the REM-off region, the orexin neurons have also been posited to project excitatory neurons to that area to prevent REM occurring during wake. In narcolepsy with cataplexy, the lack of orexin projection to the REM-off neurons weakens the inhibition of REM-on neurons, allowing for more frequent and inappropriate transitions into REM. Conversely, the eVLPO has known GABAergic projections to the REM-off vIPAG-LPT region, likely to inhibit the area during REM sleep.

HOMEOSTATIC REGULATION OF SLEEP

In 1982, Borbély proposed the classic model of sleep regulation involving both a homeostatic (Process S) and circadian (Process C) process. The sleep-dependent Process S, also known as “sleep propensity,” builds during the waking period and is dissipated by sleep. The longer an individual is wake (e.g., sleep deprivation) the higher the sleep propensity, which will require extra recovery sleep to dissipate. EEG spectral analysis-measured slow wave activity is often used as the surrogate for sleep propensity, and adenosine has been proposed as its molecular equivalent.

Adenosine was initially observed to be a possible somnogen when its injection, and injection of adenosine analogs, to the preoptic area increased total sleep primarily through an enhancement of SWS. Since then, numerous studies have demonstrated the sleep-inducing effect of adenosine, both in cats and dogs. Because of its hypnotic effect and the fact that adenosine is a by-product of energy metabolism, many postulated that adenosine is a homeostatic modulator. During wakefulness, the brain uses ATP, which eventually breaks down to adenosine and accumulates in the brain, including the basal forebrain. Indeed, basal forebrain adenosine concentration has been shown to progressively increase with each succeeding hour of sleep deprivation and decrease during SWS.

A differential effect of adenosine has been found depending of its site of activity and receptor type. In the basal forebrain, adenosine binds to A1 receptors to decrease wakefulness signals sent to the cortex. Whereas adenosine in the VLPO binds to A2 receptors to induce Fos expression and increase sleep. Most people have had personal experiences with adenosine’s role in sleepiness with the use of coffee or tea, beverages which contain the adenosine receptor antagonist caffeine or theophylline. In summary, adenosine is likely one of the factors responsible for the homeostatic regulation of sleep. More research is needed to determine other molecules that share the similar function.

CIRCADIAN REGULATION OF SLEEP

The circadian influence on sleep (process C), as proposed by Borbély, is a sleep-independent process that reflects the rhythmic variation in sleep propensity during prolonged sleep deprivation and is controlled by a circadian oscillator. In fact, circadian rhythm is exhibited by many bodily functions, including the sleep-wake cycle, body temperature, and hormone secretion (e.g., cortisol). Work performed in a forced desynchrony protocol has confirmed a strong circadian rhythm to sleep drive, especially regulation of REM sleep. Governed by
the anterior hypothalamic suprachiasmatic nucleus (SCN), the master clock of the body, human circadian rhythm cycles in a near-24-hour rhythm driven by rhythmic expressing clock genes. Because the period of the endogenous circadian rhythm is not precisely 24 hours, the master clock must entrain, or be reset, daily to entraining agents, the most powerful of which is light.

Edgar and colleagues expanded upon the model by proposing that the SCN actively initiates and maintains wakefulness and opposes the homeostatic sleep tendency during the subjective day (see Fig. 3). Much work has been performed to elucidate the link between the SCN and the sleep system. The SCN has indirect projections to both the VLPO sleep center and wake-promoting orexin neurons in the LH via the adjacent subparaventricular zone (SPZ) and the dorsomedial nucleus of the hypothalamus (DMH). The function of the SPZ was identified because, similar to SCN ablation, lesions of the ventral SPZ led to abolishment of sleep-wake circadian rhythm, along with rest–activity and feeding rhythms. The SPZ in turn projects mainly to the DMH, which has an activating effect, as lesions of this region led to more sleep and less locomotor activity. The DMH sends GABAergic neurons to the VLPO (inhibiting sleep and promote wakefulness) and glutaminergic and thyrotropin-releasing hormone neurons to the LH (excitatory to the orexin neurons). The presence of intermediate relay stations between the SCN and the sleep and wake centers of the hypothalamus allow for differential effects of the SCN in diurnal versus nocturnal animals. While the SCN is always active during the light period, its signal is alerting in diurnal organisms (wake promoting) and inhibitory in nocturnal organisms (sleep promoting). The pathways connecting the SCN and the sleep and wake centers thus provide an anatomic basis for the circadian regulation of the sleep-wake cycle.

In Borbély’s two-process model of sleep and wake, sleep propensity and duration are determined by the interaction of Processes S and C. The model is able to simulate the variations of sleep duration as a function of sleep onset time, and it describes the cyclic alternation between NREM and REM sleep as a result of reciprocal interaction between the two states. While this model has been pivotal in our conceptualization of sleep and wake, it will undoubtedly be refined as our anatomic and molecular understandings of sleep and circadian regulation evolve (Fig. 4).

**SUMMARY**

There has been significant progress in our knowledge of the mechanisms underlying the regulation of sleep and wakefulness. Novel neurotransmitters, pathways, and receptors have been discovered to refine prior theories and define new hypotheses. For example, recent studies have pointed to mutually inhibitory pathways that regulate the switch between wakefulness and sleep, and between NREM and REM sleep, much like flip-flop switches. Similarly, the importance of the wake-promoting orexin pathway has been demonstrated in animal models, as well as in humans with narcolepsy. These new findings have provided novel targets for the development of agents to manage clinical sleep disorders such as insomnia, hypersomnia, and narcolepsy. The concept of interacting circadian and homeostatic systems in regulating sleep and wakefulness has facilitated our understanding of sleep onset and maintenance. This two-process model will undoubtedly be modified as the molecular basis of each system is further elucidated. As we improve our understanding of sleep and wake regulation, novel behavioral and pharmacologic targets for the treatment of sleep disorders will emerge to improve sleep and wake functions and mitigate the negative effects of poor sleep on health.

**REFERENCES**

10. Moruzzi G, Magoun HW. Brain stem reticular forma-
24. Siegel JM, Bohemmer LN. Narcolepsy and the hypo-


75. Shouse MN, Siegel JM. Pontine regulation of REM sleep components in cats: integrity of the pedunculopontine tegmentum (PPT) is important for phasic events but unnecessary for atonia during REM sleep. Brain Res 1992;571:50–63.