Physiology of Sleep
David W. Carley1–4 and Sarah S. Farabi1,2

Why must we sleep? Pinpointing the essential and irreplaceable aspects of sleep remains one of the great challenges of mammalian biology. Still, much has been determined about the structures, processes, and pathways underlying the regulation of sleep and the relationship of sleep to daytime functioning and overall well-being. Extensive texts on these topics are available elsewhere (1,2). This article provides a concise overview of the anatomy, neurochemistry, and physiology of normal sleep and sleep homeostasis, with an eye toward the interface between sleep and metabolism. Other articles in this issue will more directly address aspects of sleep in relation to diabetes.

Generation and Maintenance of Sleep and Wakefulness

As depicted in Figure 1A, the cortical activation necessary to maintain wakefulness is supported by an extensive network of subcortical structures and pathways. Major neurochemicals of this “ascending arousal system” include excitatory norepinephrine arising from the locus ceruleus (LC), serotonin from the midline raphe nuclei, histamine from the tuberomammillary nucleus, dopamine from the ventral periaqueductal gray matter, acetylcholine from the pedunculopontine tegmentum, and the laterodorsal tegmentum of the pons and orexin from the perifornical area. Despite their apparent redundancy, normal behavioral functioning may require all of these arousing systems. For example, it is now clear that narcolepsy results from a selective loss of orexin-releasing neurons in the forebrain, accounting for the excessive daytime sleepiness, fragmented sleep, and cataplexy (sudden muscle weakness without loss of consciousness) associated with this disorder.

Initiation and maintenance of sleep require suppression of activity in the ascending arousal systems. This is accomplished by inhibitory neurons of the ventrolateral preoptic area (VLPO; Figure 1B), which remain active throughout sleep (3). The molecular “triggers” that activate the VLPO and initiate sleep onset have not been fully defined, but a substantial body of evidence points to extracellular adenosine as a candidate. Adenosine accumulates in basal forebrain during wakefulness and diminishes with ongoing sleep (4). Adenosine receptors are expressed in the VLPO and adenosine activates VLPO neurons in vivo (5), making it a reasonable candidate for the “sleep
switch.” Caffeine and theophylline are potent adenosine receptor antagonists, which may form the basis for their well-known alerting effects. Despite this evidence, it is almost certain that other molecules also play important signaling roles controlling the initiation and maintenance of sleep. The monoaminergic arousal centers project to the VLPO and may serve to inhibit its activity (6). This creates the concept of “flip-flop” control of behavioral state, in which, at any given time, activity of either arousal-producing or sleep-producing neurons dominates and suppresses the other (3). In addition, the VLPO receives important circadian modulation from the suprachiasmatic nucleus—the central circadian clock (3).

Sleep itself is not a homogenous process. There exist two fundamentally distinct types of sleep: rapid eye movement (REM) sleep, which is associated with active dreaming, and non–rapid eye movement (NREM) sleep. Switches between NREM and REM sleep appear to be controlled by reciprocal inhibition between monoaminergic neurons and a specific subset of cholinergic neurons within the brainstem (7). These “REM-on” cholinergic neurons exhibit reciprocal inhibitory connections to noradrenergic (LC) and serotonergic (raphe) neurons (8). When REM sleep is triggered, REM-on cholinergic neurons become maximally active, while noradrenergic and serotonergic neurons become virtually silent. The switching between activity and inhibition of these neurons results in a characteristic cycling between NREM and REM during the sleep period.

**Measurement and Quantification of Sleep and Wake States**

Assessment of sleep/wake states can be made by behavioral observation, physiological monitoring, or a combination of the two. Behaviorally, sleep in adults is characterized by loss of consciousness and by relative immobility in a recumbent posture with the eyes closed. During NREM sleep, there is reduced tonus of large skeletal muscles that progresses to complete or near-complete atonia with a transition to REM sleep. Throughout sleep, there is a relative sparing of activity among respiratory pump muscles. Visual, olfactory, auditory, somatosensory, and even nociceptive sensory responses all are diminished but not eliminated during sleep (9). Furthermore, many sensory responses exhibit differing characteristics during NREM versus REM sleep.

Physiologically, the gold standard for assessment of sleep and wake states is the laboratory polysomnogram (PSG). To conduct a PSG, numerous noninvasive sensors are attached to a subject. These sensors include multiple skin electrodes, which record brain activity (electroencephalogram [EEG]), eye movements, submental muscle tone, leg movements, and electrocardiogram (ECG). Thoracic and abdominal strain gauges, oral and nasal airflow sensors, and a finger probe to measure arterial oxygen saturation are also attached to the subject to help monitor respiration during sleep.

In addition to wakefulness and REM sleep, current clinical guidelines for scoring PSGs identify three stages of progressively deepening NREM sleep: stages N1–N3 (10). These stages are recognized and scored based on characteristic rhythms and events observed in the PSG waveforms, but a detailed presentation of the scoring process is beyond the scope of this article (11). Briefly, alert wakefulness is associated with a low-amplitude mixed frequency EEG pattern. As illustrated in Figure 2, drowsy wakefulness is associated with alpha waves seen as a rhythm with peaks in the 8- to 13-Hz range. Drowsiness also is associated with slow rolling eye movements that may persist into light sleep. The lightest stage of NREM sleep (N1) is characterized by a loss of alpha rhythm and presence of theta waves with a characteristic frequency of 4–7 Hz. Stage N2 sleep is marked by the expression of spindles (burst-like trains of waves in the 11- to 16-Hz range with a total duration ≥0.5 seconds) and K-complexes (well-defined biphasic waves lasting ≥0.5 seconds and usually maximal over the frontal cortex). Deep NREM sleep (stage
N3) is associated with large (≥75 µV) slow (0.5–3 Hz) waves known as delta waves. Typically, skeletal muscle activity exhibits progressively decreasing amplitude with transitions from wakefulness to N1, N2, and N3 sleep. REM sleep is associated with the lowest skeletal muscle tone and with either sharp theta waves (sawtooth waves) or wake-like EEG patterns (Figure 2).

For scoring purposes, an overnight PSG recording is divided into 30-second epochs, and a stage score is assigned to each epoch. Visualizing this sequence of stage scores graphically as a “hypnogram” highlights the temporal structure of the sleep process (Figure 2). During a normal night, the sleep process is cyclical, with sleep onset being followed by a rapid descent to deep stage N3 sleep within the first hour. This is followed by cyclical alternations between NREM and REM sleep occurring every 60–90 minutes throughout the rest of the night. Typically, most N3 sleep occurs during the first half-night, whereas most REM sleep occurs during the second half-night (Figure 2). The full biological and clinical relevance of this “ultradian” cycling of sleep depth remains to be determined. PSG data also are amenable to quantitative and continuous analysis using various signal processing techniques. Because the EEG rhythms associated with different “levels” of alertness and NREM sleep can be differentiated according to characteristic frequencies, EEG power spectrum analysis has become a very popular sleep research tool (12).

Sleep and sleep quality also can be assessed subjectively, by self-report of the individual patient or research subject. Various mobile phone application–based, Web-based, or paper sleep and activity logs have been developed and can be useful adjuncts to laboratory or home testing for both clinical and research purposes. Although numerous standardized survey instruments have been developed to assess sleep quality, two have been very widely used: the Pittsburgh Sleep Quality Index (13) and the Functional Outcomes of Sleep Questionnaire (14).

In-home PSG testing has been performed without monitoring (15), with remote monitoring (16), and with in-home monitoring by a trained technician (17). The frequency of technically unacceptable recordings using these approaches ranges from about 10% to about 30%, and it is unclear whether any consistent cost savings accrue. Numerous other methods for home sleep testing have been developed that do not incorporate EEG monitoring. The simplest of these is actigraphy, in which the subject wears a device—typically the size and shape of a wristwatch—on the nondominant wrist. The device contains multiple accelerometers and can record movements continuously for periods of up to several weeks. These movement profile data are then used to discriminate sleeping and waking periods, and their overall agreement with PSG-based determinations is good, although results of individual validation studies have varied (18). This approach has been popular in research studies, allowing collection of many days of sleep/wake patterns in a natural environment. Some actigraphic devices also contain light intensity meters and eventlogging buttons, allowing subjects to note when they go to bed and when they arise. For clinical applications, “level-3” cardiorespiratory monitoring devices based on a wide array of technologies increasingly are being employed, but a thorough review is beyond the scope of this article. Typically, these devices monitor heart rate and its variability, respiratory effort and airflow, and arterial oxygen saturation. These systems are commonly used to screen for clinically significant sleep apnea syndrome (19).

**Endocrine Manifestations of Sleep and Wake States**

Plasma levels of most hormones exhibit significant 24-hour rhythms (20,21), pointing to the importance of both the circadian clock and sleep-specific influences on their release and/or metabolism. Some hormones are little influenced by sleep versus wakefulness, including adrenocorticotrophic hormone, cortisol, and melatonin; some are strongly influenced by sleep, such as thyroid-stimulating hormone (TSH) and prolactin; and some are affected by particular sleep stages, such as growth hormone (20).

Under normal conditions, prolactin levels are low during the daytime and high during sleep at night. Studies using daytime naps or sudden changes in sleep schedule have shown that sleep onset, regardless of time of
sleep, appears to exert an inhibitory influence on TSH secretion (20). N3 sleep is consistently associated with falling TSH levels, whereas awakenings are associated with rising TSH levels (25). This may be significant in diabetes pathogenesis, as TSH level has been negatively associated with various measures of insulin sensitivity. However, cause-and-effect relationships remain to be determined (26).

**Conclusion**

In summary, sleep is an actively regulated process significantly modulated by homeostatic influences that accumulate during ongoing wakefulness and dissipate during sleep and by circadian effects entrained to the 24-hour day. Sleep has a typical underlying architecture characterized by a rhythmic alternation between NREM and REM stages, and the transitions among sleep/wake states are orchestrated by a well-defined subcortical network of brain structures. Sleep and wake states also are characterized by distinct hormonal patterns that exert potential significant influences on metabolism and glucose homeostasis.

**Duality of Interest**

No potential conflicts of interest relevant to this article were reported.

**References**