

Neurobiology of Sleep and Wakefulness

Dr. Robert W. McCarley, Harvard Medical School, Boston, MA

Dr. Christopher M Sinton, University of Texas Southwestern Medical Center, Dallas, TX

Sleep and wakefulness are brain states that depend on specific systems of neurons for their onset maintenance, and termination.

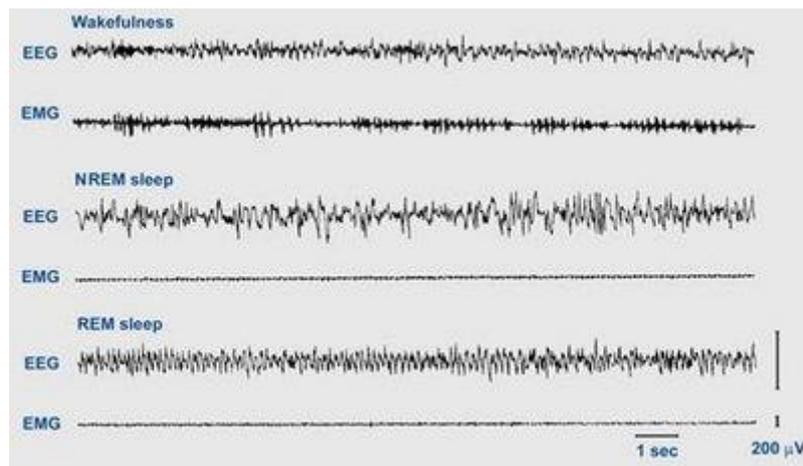


Figure 1: Electroencephalogram (EEG) and electromyogram (EMG) signals can be used to define the states of wakefulness and sleep. This example is from a rodent, but all mammals show these states.

Sleep is a natural, periodically recurring state of inactivity, characterized by the loss of consciousness and reduced responsiveness to external stimuli. In contrast, wakefulness is the absence of sleep and is marked by consciousness, awareness and activity. Observation and speculation about the state of sleep was possible for thousands of years. However, it was only in 1929 that the German psychiatrist, Hans Berger discovered that the electrical activity of the brain could be recorded as brain waves and that these waves changed as wakefulness gave way to sleep. This discovery of the electroencephalogram (EEG), as it is now known, made the objective study of sleep possible. By detecting synchronous activity of cortical neurons and recording voltage fluctuations in terms of the amplitude of the resulting waves and their frequency, the EEG is thus used to differentiate changes in alertness and sleep stages. For the analysis of the states of sleep and wakefulness, EEG frequencies are conveniently grouped into bands:

- Delta, 0.5 to 4 Hz
- Theta, 4 to 8 Hz
- Alpha, 8 to 12 Hz

- Sigma, 12 to 14 Hz
- Beta, 14 to 30 Hz
- Gamma, 30 to 50 Hz

EEG data are combined with those from concurrent recording of eye movements from the electrooculogram (EOG), and muscle tone from the electromyogram (EMG) to define the states of sleep and wakefulness. Other signals, such as breathing rate, are also recorded during polysomnographic evaluation of human sleep in the clinic but they are not required for state definition. In animals, EEG and EMG signals can be sufficient to define the states of sleep and wakefulness.

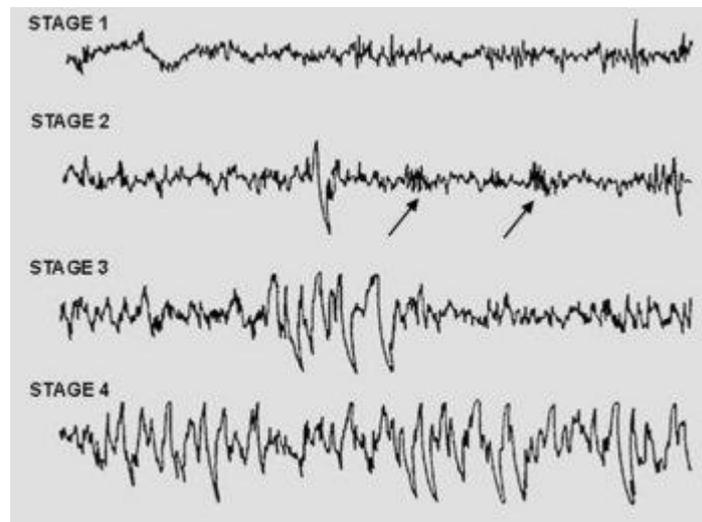


Figure 2: In the human, the gradual slowing of the EEG, and the increasing presence of delta activity, enables NREM sleep to be further classified into four stages. The presence of spindles is marked by the arrows during Stage 2.

Wakefulness, as shown in Figure 1, is defined by a low voltage fast frequency EEG pattern, called desynchronized or activated EEG, that consists primarily of frequencies in the beta and gamma ranges. The EOG and EMG recordings also show high activity during wakefulness.

In contrast to wakefulness, sleep is characterized by higher voltages and slower waves, a pattern called synchronized EEG. Non-rapid eye movement (NREM) sleep is the usual term for this state. In humans, NREM sleep is further subdivided into four Stages, 1 through 4, which depend on the extent of EEG slowing, especially in the delta frequency range (Figure 2). Stage 2 is notable for the presence of spindles, which are waxing and waning bursts of frequencies in the sigma band. Stages 3 and 4, with delta waves comprising more than 50% of the signal, are further grouped under the term slow wave sleep (SWS). During this state, the EOG can show

gradual rolling eye movements and there is low or minimal muscle activity. In animal studies, the terms SWS and NREM sleep are often used synonymously as the state is rarely subdivided in terms of EEG slowing.

Slow sleep is interrupted by periods of rapid eye movement (REM, i.e., active or paradoxical) sleep, when, despite all the overt signs of continuing sleep, the activity of the brain is remarkably different. In fact, the EEG in humans during REM sleep is essentially identical to that recorded during wakefulness, but the EOG reveals rapid bursts of eye movements, hence the name of the state. In rodents, the EEG is typically dominated by frequencies in the theta band during REM sleep and phasic activity appears as twitching of the vibrissae as well as eye movements. Importantly, and in all species, the EMG shows the complete loss of muscle tone (i.e., atonia) that is a characteristic of REM sleep.

The natural progression of the states of wakefulness and sleep

At sleep onset in humans, the low voltage, high frequency EEG pattern of wakefulness, often with alpha waves when the eyes are closed, gradually changes to Stage 1 sleep as the EEG frequencies slow. Stage 1, a brief transitional phase after wakefulness, is followed by Stage 2 sleep when EEG frequencies slow further. During Stage 2, the episodic bursts of rhythmic, 12- to 14-Hz waveforms occur in the EEG. These bursts are sleep spindles, an important characteristic of sleep onset. Stage 2 is followed by Stages 3 and 4, which are defined by the presence of increasing amounts of high amplitude, slow delta waves. To complete the sleep cycle, REM sleep begins when a low voltage, high frequency EEG pattern returns, but accompanied by rapid eye movements and muscle atonia.

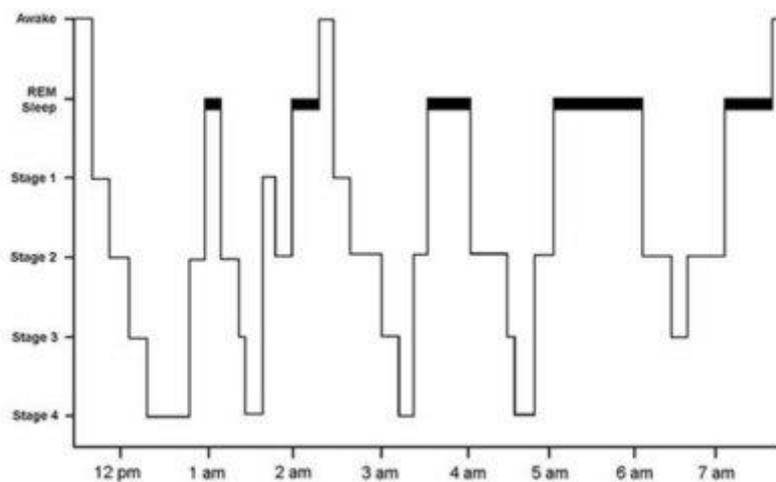


Figure 3: A typical night of sleep in a young human adult is displayed in this hypnogram.

Sleep typically and predictably proceeds from one stage to the next (Carskadon and Dement, 2005). As sleep progresses, a stepwise descent from wakefulness to Stage 1 through to Stage 4 sleep occurs, followed by an abrupt ascent back through a Stage 2-like EEG with spindles towards Stage 1 with irregular theta activity (Figure 3). However, at Stage 1, the first REM sleep episode usually occurs, about 70 to 90 minutes after sleep onset in the human. Following the first REM sleep episode, the sleep cycle repeats with the appearance of NREM sleep. Then, about 90 minutes after the start of the first REM sleep period, another episode occurs. This rhythmic cycling persists throughout the night. Hence the sleep cycle duration is about 90 minutes in humans, and, on average, REM sleep episodes last for about 20 minutes, gradually increasing in duration throughout the night. Over the course of the night, delta wave activity declines, and as sleep progresses, the EEG of NREM sleep is comprised increasingly of waves of higher frequencies and lower amplitude.

Wakefulness and the desynchronized EEG

Wakefulness depends on forebrain activation

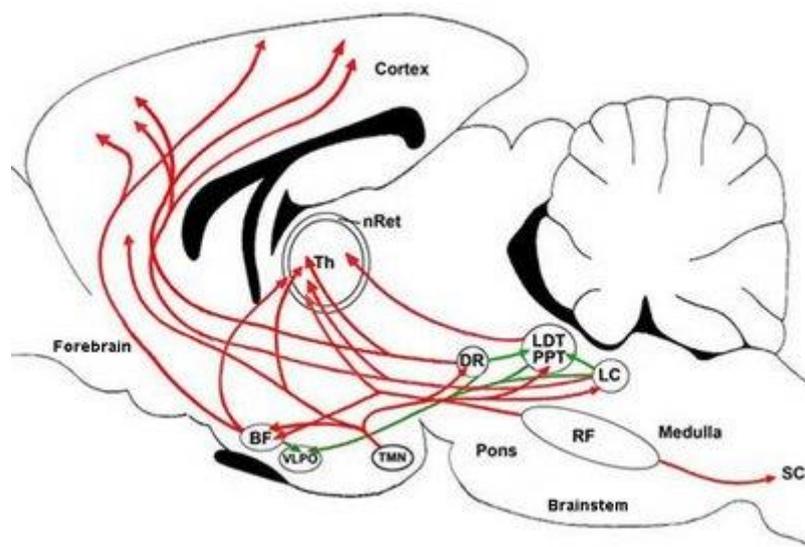


Figure 4: The principal activating systems that support wakefulness are displayed on this schematic of a sagittal section of a rodent brain. Excitatory influences are shown in red and inhibitory in green. See text for abbreviations.

Wakefulness and EEG desynchronization require excitatory innervation of the forebrain. Being awake depends on the discharge activity in several, apparently redundant parallel ascending neurotransmitter pathways: none of these systems, which include glutamate, acetylcholine (ACh), and the monoamines (serotonin and norepinephrine), is absolutely necessary for the expression of wakefulness, but all appear to contribute (Jones, 2005). These

pathways are shown in Figure 4. Many of the corresponding cell bodies for these systems are located in the brainstem. Two studies were critical for the discovery that this region was a principal substrate for wakefulness and hence gave rise to the important concept that wakefulness entails forebrain activation for its expression. Bremer, in a 1935 publication, noted that a complete transection above the brainstem produces a preparation that remains in a state resembling normal sleep. Subsequently, Moruzzi and Magoun in 1949 demonstrated that electrical stimulation of the reticular core of the upper brainstem produced immediate and long-lasting EEG desynchronization in a previously sleeping preparation. The latter study thus established the region as the location of an *ascending reticular activating system* (ARAS) that maintained wakefulness although the neurons that were responsible remained unknown.

Excitatory neurotransmission supports wakefulness

The neurotransmitter systems that comprise the ARAS were subsequently identified. However, when these transmitter systems were described individually, the term ARAS became less informative, especially as some of the cell bodies were found to lie outside the core of the reticular formation (RF). The *ascending activating system* (AAS) is a general term that is now more typically used. A major component of the AAS was determined by Steriade and collaborators (Steriade et al., 1990), who recorded the activity of neurons in chronic, freely moving cats. They identified thalamically projecting cells located at the pons-midbrain junction whose discharge rates increased before the first sign of the change to an EEG desynchronized state. Later studies established that many of these cells which project to the thalamus (Th) were cholinergic and localized them to the laterodorsal tegmentum/pedunculopontine (LDT/PPT) region. They remain active whenever the EEG is desynchronized during wakefulness and REM sleep. Conversely these neurons have low discharge rates during NREM sleep.

The LDT/PPT cholinergic system is not the exclusive substrate of EEG desynchronization. Other significant brainstem reticular neuronal projections to the thalamus, utilizing glutamatergic neurotransmission, as well as monoaminergic projections that contain norepinephrine from the locus coeruleus (LC) and serotonin from the dorsal raphe (DR) also play a role and are part of the AAS (McCormick, 1989). The monoaminergic cells send widespread projections throughout the forebrain to the cortex as well as the thalamus. Importantly, they demonstrate significant state-related changes in discharge rates, being highest during wakefulness, decreasing during NREM sleep and becoming virtually quiescent during REM sleep. A similar state-related change in discharge rate is also shown by cells that contain another monoamine, histamine. These cells are located uniquely in the tuberomammillary nucleus (TMN) of the caudolateral hypothalamus. Histaminergic cells, like

the other monoaminergic components of the AAS, have been found to be a substrate of wakefulness on the basis of pharmacological, electrophysiological, lesion and stimulation studies. A second ascending cholinergic projection, originating in the nucleus basalis of Meynert in the basal forebrain (BF), sends widespread projections to the cortex and thalamus. This cholinergic system, like that from the LDT/PPT, is also important for EEG desynchronization, is a component of the AAS, and receives brainstem input. Thus, there are two major relay pathways for brainstem activation of cortex, one through the thalamus and one through the BF and nearby hypothalamus, as well as the direct projections of LC and DR to cortex.

Another relatively recent addition to the list of neurotransmitters that contribute to wakefulness is the neuropeptide, orexin (also called hypocretin). Orexin-containing cell bodies are located uniquely in the lateral hypothalamus and send widespread projections throughout the neuraxis. The importance of orexin to the control of sleep and wakefulness was discovered serendipitously when narcolepsy was found to be caused by the absence of an orexin signal, either through the loss of the receptors, the neurons or the neuropeptide. Orexin supports wakefulness through excitatory projections to other components of the AAS, though its role in the regulation of sleep and wakefulness is likely to be more critical than the enhancement of general arousal and EEG desynchronization. One significant conclusion, now supported by considerable research, is that orexin is important for controlling the gate from one state to another (Lu et al., 2006; Saper et al., 2001). Hence orexin ensures the orderly transition between states and also the rapid switching that is evident when a state transition is initiated.

NREM sleep and EEG desynchronization

Sleepiness

Sleepiness is determined by two factors. One, a *homeostatic* factor, depends on the duration since the last sleep bout. The longer the time spent awake, the greater the drive for sleep. Evidence points to adenosine, a neurotransmitter with extracellular levels coupled to cellular metabolism as one mediator of the homeostatic sleep response. Results show that adenosine accumulates during wakefulness and can inhibit components of the AAS, in particular BF cholinergic and non-cholinergic neurons (McCarley, 2007). This would make it a potential homeostatic sleep factor. The second factor is *circadian*, which varies with a 24 hour periodicity and is independent of the amount of preceding sleep or wakefulness. The circadian influence is modulated through the suprachiasmatic nucleus (SCN) with inhibitory projections to components of the AAS. The discharge activity of cells in the SCN depends on the circadian

cycle (Inouye and Kawamura, 1979). Together, the homeostatic and circadian factors modulate the need for sleep and influence the balance between alertness and sleepiness.

A sleep-promoting area

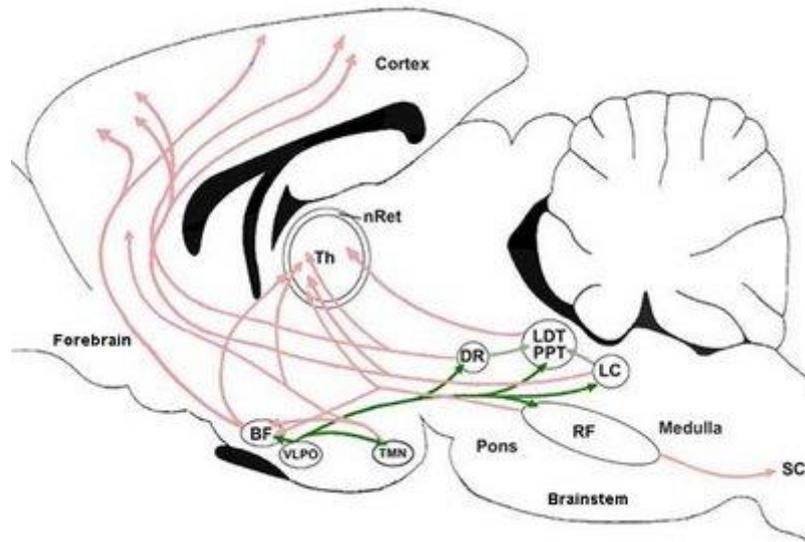


Figure 5: NREM sleep is marked by a reduced discharge rate of the activating systems, shown here in pink, as the inhibitory influence from cells in the VLPO region increases.

The existence of a sleep-promoting area in the brain was first implied from observations of the neuropathology of encephalitis by von Economo, who published a correlation between damage to the anterior hypothalamus and profound insomnia in 1930. The importance of this region was later confirmed by stimulation, electrophysiological and histological studies, which identified the critical area as being within the hypothalamic preoptic area of the BF. In particular, a group of cells that are concentrated in the ventrolateral preoptic area (VLPO) (Figure 5) has been shown to be active specifically during sleep and to contain the inhibitory neurotransmitters, γ -amino butyric acid (GABA) and galanin. VLPO cells show their first signs of activity during drowsiness and continue to discharge selectively throughout NREM sleep (Szymusiak, 1995). They are inhibited by the neurotransmitters of the AAS and also send direct inhibitory projections to all components of the AAS. Hence the neurons of the VLPO and AAS disinhibit their own influence, i.e. their activity is self-reinforcing. Such a bistable system ensures not only the stability of wakefulness or NREM sleep, but also the rapid transition between the states when the balance is altered by homeostatic and/or circadian factors once a certain threshold is reached. This minimizes prolonged intermediate episodes of drowsiness

and half-sleep or half-awake states. Orexin could also play a role in this mechanism, for example by coordinating the activity of different components of the AAS.

The neurobiology of REM sleep

REM sleep is generated in the brainstem

After a series of transection studies, Jouvet and collaborators localized the neurons required for the generation of REM sleep to the brainstem (Jouvet, 1979). A transection made just above the junction of the pons and midbrain produced a state in which the periodic occurrence of REM sleep signs could be found in the isolated brainstem while, in contrast, recordings in the isolated forebrain showed no sign of REM sleep.

Having established the importance of the brainstem in REM sleep, subsequent work has provided a more detailed picture of the circuits controlling the REM sleep rhythm. As in humans, the cardinal signs of REM sleep in all species in which the state occurs include EEG desynchronization, rapid eye movements and muscle atonia. Importantly, these components can be dissociated, indicating that each is under the mechanistic control of a different group of neurons, which have been found to be especially concentrated in the pontine RF. These cell groups have been termed effector neurons (McCarley, 2007; Steriade and McCarley, 2005).

REM sleep effector neurons

Intracellular recordings of effector neurons have shown that they remain in a relatively hyperpolarized state and generate almost no action potentials during NREM sleep, but that they begin to depolarize even before the occurrence of the first sign of EEG desynchronization that signals the approach of REM sleep. Then, as depolarization of the pontine RF neurons proceeds and the threshold for action potential production is reached, they begin to discharge. A high rate of discharge is maintained throughout the REM sleep episode, due to continuing depolarization.

These different groups of effector neurons that control the components of REM sleep are centered in the pontine RF. This is an anatomically complex region and hence detailed separation and demarcation between the groups is only approximate:

- Eye movements are controlled by a group of medial pontine RF neurons;
- Muscle atonia by a group of neurons in the lateral pontine RF, including the ventrolateral and dorsolateral pontine RF. These send inhibitory projections to the motor neurons in the spinal cord (SC);

- EEG desynchronization by neurons in an extensive pontomesencephalic field, which includes the LDT/PPT cholinergic neurons. Additionally, BF cholinergic neurons affect EEG desynchronization.

REM-on and REM-off cells

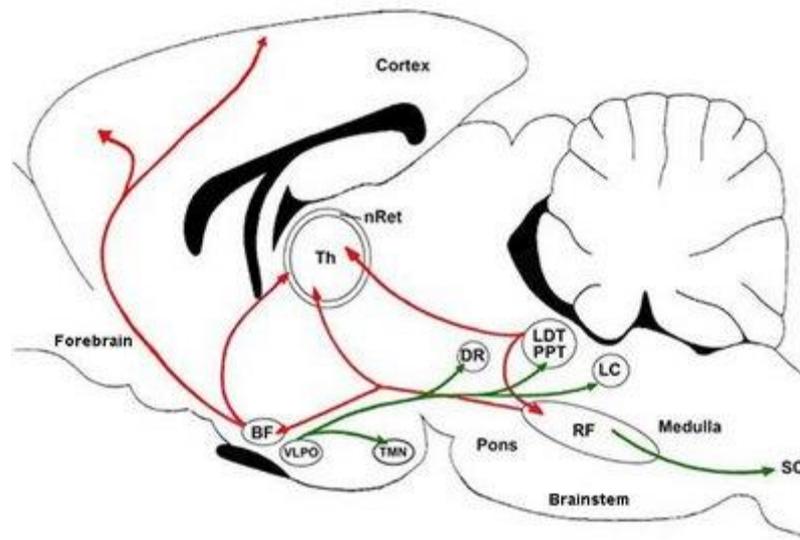


Figure 6: REM sleep is characterized by inhibition of the monoaminergic (LC and DR) activating systems. Hence the REM-on cells of the LDT/PPT, which project caudally to the RF, are disinhibited. Forebrain activation is supported by cholinergic systems (LDT/PPT and BF) during REM sleep.

A subset of the brainstem LDT/PPT neurons are not active during wakefulness but discharge preferentially just before and during REM sleep: these cells are likely to be cholinergic (el Mansari et al., 1989; Thakkar et al., 1998) (Figure 6). Furthermore, lesions and electrical stimulation of the LDT/PPT produce, respectively, a decrease and an increase in REM sleep, and direct in vivo measurements of ACh demonstrate that pontine levels of the neurotransmitter are higher during REM sleep. This evidence leads to the conclusion that these LDT/PPT neurons act as promoters of REM sleep signs via their excitatory projections to the various populations of effector cells in the PRF. Hence they are termed REM-on cells.

In contrast, the nearby serotonergic and noradrenergic neurons are REM-off cells and suppress REM sleep via inhibitory projections specifically to the REM-on subset of LDT/PPT cholinergic neurons (McCarley, 2007; Thakkar et al., 1998). Monoaminergic cells exhibit a pattern of discharge activity that is nearly opposite to that of LDT/PPT cholinergic cells: their discharge rate is greatest during waking, declines during NREM sleep and virtually ceases prior

to and during REM sleep. The TMN histaminergic neurons exhibit the same change in discharge rate but are not essential for REM sleep cyclicity.

In this model, as the discharge activity of brainstem monoaminergic neurons decreases with NREM sleep, less monoamine neurotransmitter is released onto the inhibitory receptors on the REM-on LDT/PPT cholinergic neurons, causing them to be disinhibited. This in turn leads to increased discharge of the LDT/PPT neurons, which initiate REM sleep primarily via their projections to the REM sleep effector neurons in the PRF.

The complexity of brainstem neuronal circuits involved in the control of REM sleep is probably not yet fully appreciated. For example, although pontine GABAergic neurons have been critically implicated in REM sleep mechanisms for some time, data were recently presented indicating that mutually inhibitory REM-on and REM-off localized populations of these cells act in concert to gate the appearance of REM sleep (cf. Lu et al., 2006 for these results, including the earlier GABA studies). And McCarley (2007) has extended the model described above that is based on cholinergic-monoaminergic interaction by incorporating GABAergic inhibitory influences.

The cellular basis of the brain state and EEG

The thalamus (Th) drives network oscillations via cortical projections

The thalamus plays a major role in orchestrating the change in the discharge pattern of cortical neurons that underlies the EEG differences between wakefulness and NREM sleep. This is because thalamic neurons, like cortical pyramidal neurons, have intrinsic membrane properties that cause their discharge pattern to change as a function of the level of depolarization of the cell. When depolarized these neurons discharge in a single-spike mode, but when hyperpolarized they display a bursting pattern. Since their level of depolarization is dependent on AAS projections to the forebrain, the discharge patterns of thalamic cells will thus be modulated by the varying levels of the AAS neurotransmitter levels, as wakefulness alternates with the states of sleep.

Spindle waves and increasing slow wave delta activity

As sleep begins and progresses into Stage 2, spindle pacemaker GABAergic thalamic nucleus reticularis (nRet) neurons, which are hyperpolarized during wakefulness, gradually depolarize as AAS input decreases. Spindle waves then arise as a discharge pattern caused by network interactions between these spindle pacemaker neurons and thalamocortical projection neurons (Steriade and McCarley, 2005). As sleep progressively deepens, spindle waves are blocked by continuing changes in AAS input to the thalamus. These changes affect the level of

depolarization of thalamic spindle pacemaker neurons as well as thalamocortical projection neurons. At this time, NREM sleep becomes increasingly characterized by high voltage, slow wave delta activity in the cortex. The cellular basis of this activity depends on thalamocortical neurons, maintained in a hyperpolarized state by the absence of depolarizing input and generating synchronous bursts of discharges (McCormick and Bal, 1997; Steriade et al., 1993). Delta waves reflect these bursts of activity, transferred through the thalamocortical network, and synchronized as oscillations with cortical pyramidal cells, which are themselves discharging in a similar mode. Importantly, this burst discharge mode in thalamic cells prevents the transfer of sensory information through the thalamus to the cortex, so maintaining the block on sensory input that is characteristic of sleep.

As noted above, another major influence on cortical activation originates in the BF, an important site for sleep homeostatic control and whose inactivation leads to increased slow wave activity. Conversely, increased bursting activity of BF cholinergic neurons is associated with cortical activation and the appearance of gamma and theta frequencies on the EEG (Lee et al., 2005).

EEG activation and the block of delta waves

Neurotransmitters that depolarize thalamocortical and cortical cells will therefore prevent the appearance of delta waves. During wakefulness, components of the AAS, including cholinergic BF and LDT/PPT plus brainstem noradrenergic, serotonergic, and glutamatergic projections, maintain this depolarized state and are thus critical for the suppression of slow wave activity. At this time, depolarized thalamocortical relay cells show the single-spike discharge mode, which allows the transfer of sensory input through the thalamus to the cortex. During REM sleep, in contrast, the cholinergic and glutamatergic influence provide the depolarizing input without a contribution from the monoaminergic cells which are quiescent.

References

Carskadon, M.A. and Dement, W.C. (2005). Normal human sleep: an overview. In: Principles and Practice of Sleep Medicine, 4th edition. M.H. Kryger, T. Roth, W.C. Dement, eds. Elsevier Saunders, Philadelphia, PA, pp. 13-23.

el Mansari, M., Sakai, K., Jouvet, M. (1989). Unitary characteristics of presumptive cholinergic tegmental neurons during the sleep-waking cycle in freely moving cats. *Exp Brain Res.* 76, 519-529.

- Inouye, S.T. and Kawamura, H (1979). Persistence of circadian rhythmicity in a mammalian hypothalamic "island" containing the suprachiasmatic nucleus. *Proc Natl Acad Sci U S A*, 76:5962-5966.
- Jones, B.E. (2005). From waking to sleeping: neuronal and chemical substrates. *Trends Pharmacol Sci*, 26:578-586.
- Jouvet, M. (1979). What does a cat dream about? *Trends Neurosci*, 2:280-282.
- Lee, M.G., Hassani, O.K., Alonso, A., Jones, B.E. (2005). Cholinergic basal forebrain neurons burst with theta during waking and paradoxical sleep. *J Neurosci*, 25, 4365-4369.
- Lu, J., Sherman, D., Devor, M. and Saper, C.B. (2006). A putative flip-flop switch for control of REM sleep. *Nature*, 441:589-594.
- McCarley, R.W. (2007). Neurobiology of REM and NREM sleep. *Sleep Med*, 8:302-330.
- McCormick, D.A. (1989). Cholinergic and noradrenergic modulation of thalamocortical processing. *Trends Neurosci*, 12:215-221.
- McCormick, D.A. and Bal, T. (1997). Sleep and arousal: thalamocortical mechanisms. *Annu Rev Neurosci*, 20:185-215.
- Saper, C.B., Chou, T.C. and Scammell, T.E. (2001). The sleep switch: hypothalamic control of sleep and wakefulness. *Trends Neurosci*, 24:726-731.
- Steriade, M., Datta, S., Pare, D., Oakson, G. and Curro Dossi, R.C. (1990). Neuronal activities in brain-stem cholinergic nuclei related to tonic activation processes in thalamocortical systems. *J Neurosci*, 10:2541-2559.
- Steriade, M., McCormick, D.A. and Sejnowski, T.J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, 262:679-685.
- Steriade, M. and McCarley, R.W. (2005). *Brain control of wakefulness and sleep*. Kluwer Academic/Plenum, New York, NY.
- Szymusiak, R. (1995). Magnocellular nuclei of the basal forebrain: substrates of sleep and arousal regulation. *Sleep*, 18:478-500.
- Thakkar, M.M., Strecker, R.E. and McCarley, R.W. (1998). Behavioral state control through differential serotonergic inhibition in the mesopontine cholinergic nuclei: a simultaneous unit recording and microdialysis study. *J Neurosci*, 18:5490-5497.