Science In The Cloud

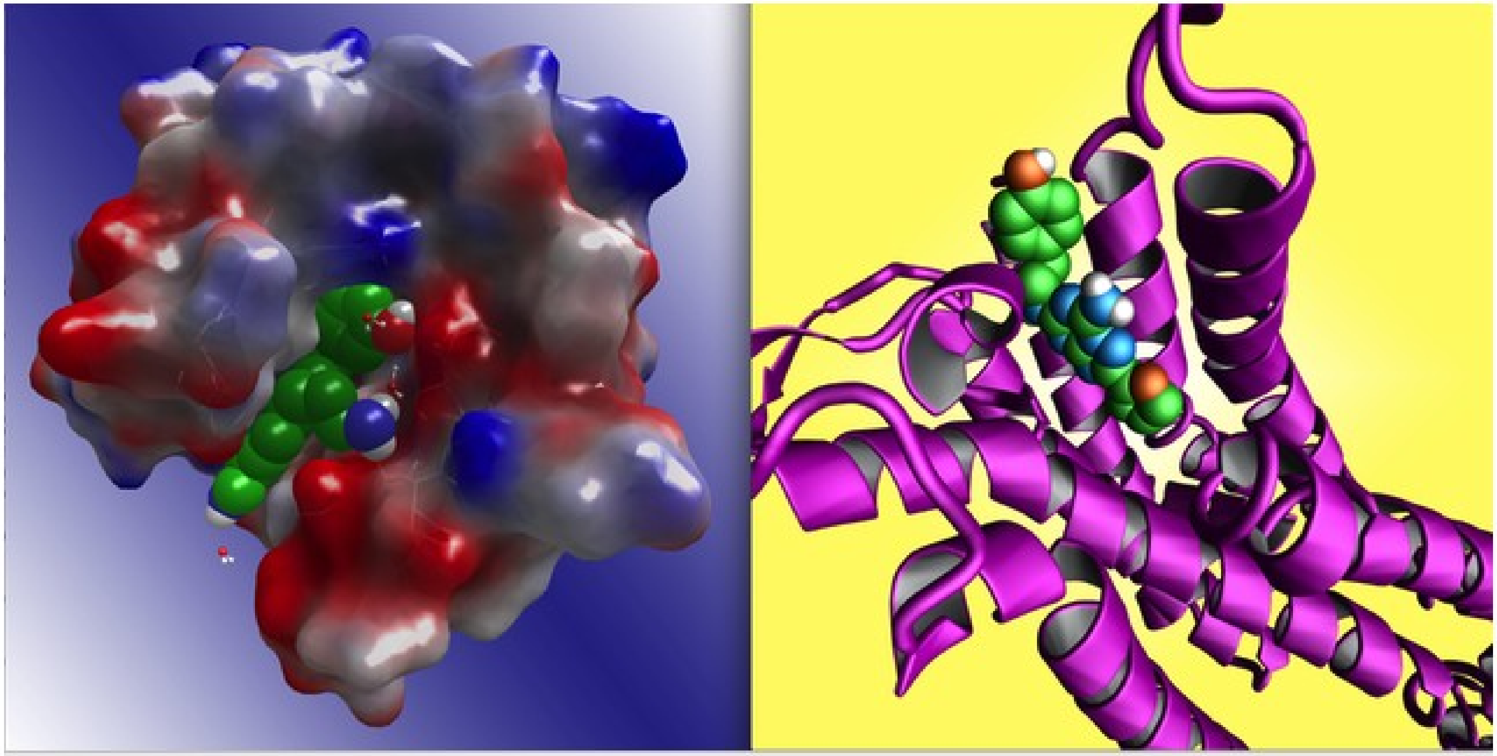
Joseph L. Hellerstein
Computational Discovery Department
Google
Seattle, Washington USA
May 7, 2012

Agenda

- Supercomputing in the cloud
  - Application to computational chemistry
- Computational discovery for brain science
  - Brain science 101
  - Simulating sleep-wake cycles
- Lessons learned, random musings

$4,829-per-hour supercomputer built on Amazon cloud to fuel cancer research

By Jon Brodkin | Published 2 days ago



Simulated images of compounds studied in pharmaceutical research

~50K cores
~6.7K EC2 instances
$4,828.85 per hour

Exacycle

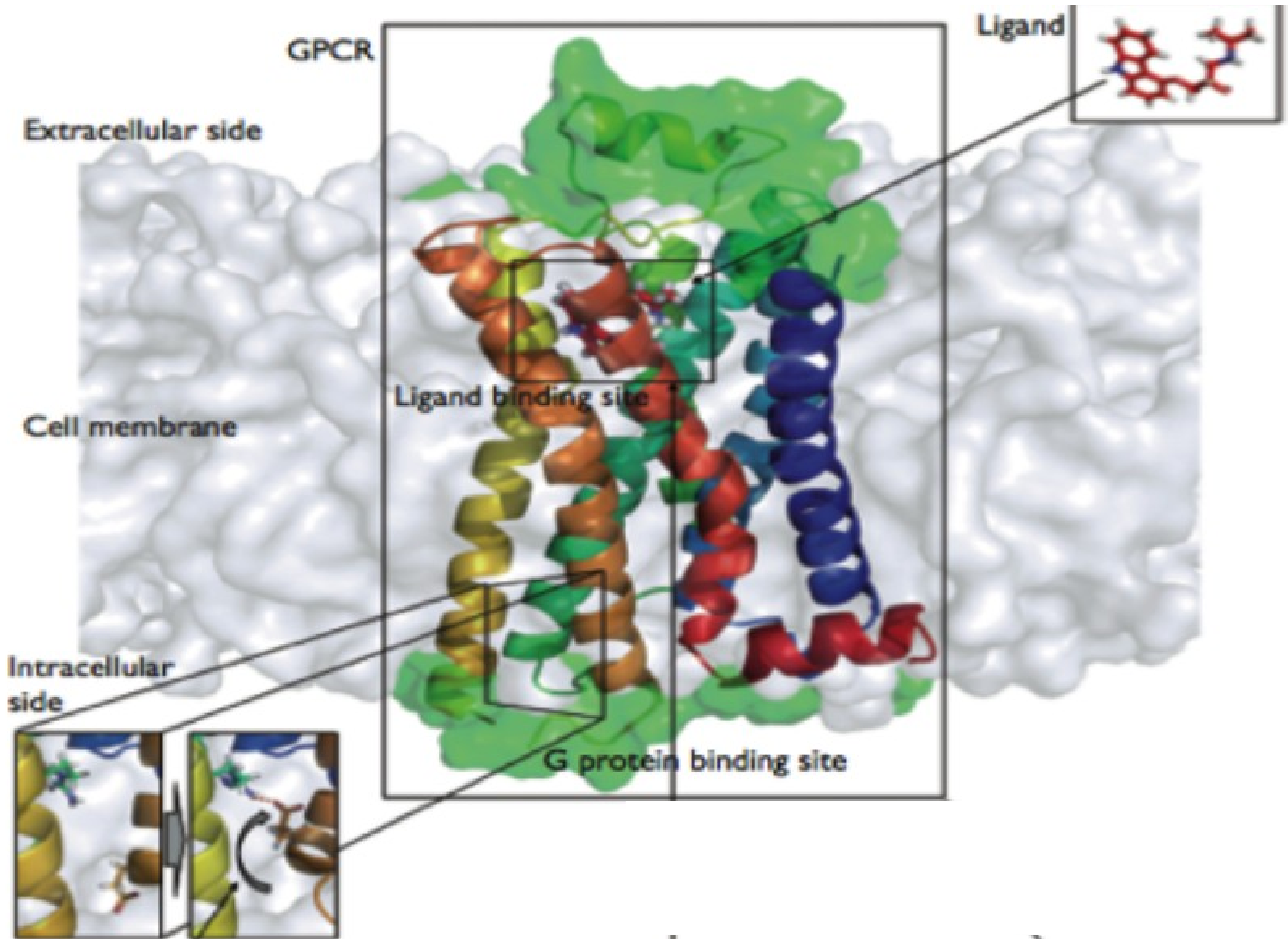
APR
7

1 billion core-hours of computational capacity for researchers

Posted by Dan Belov, Principal Engineer and David Konerding, Software Engineer

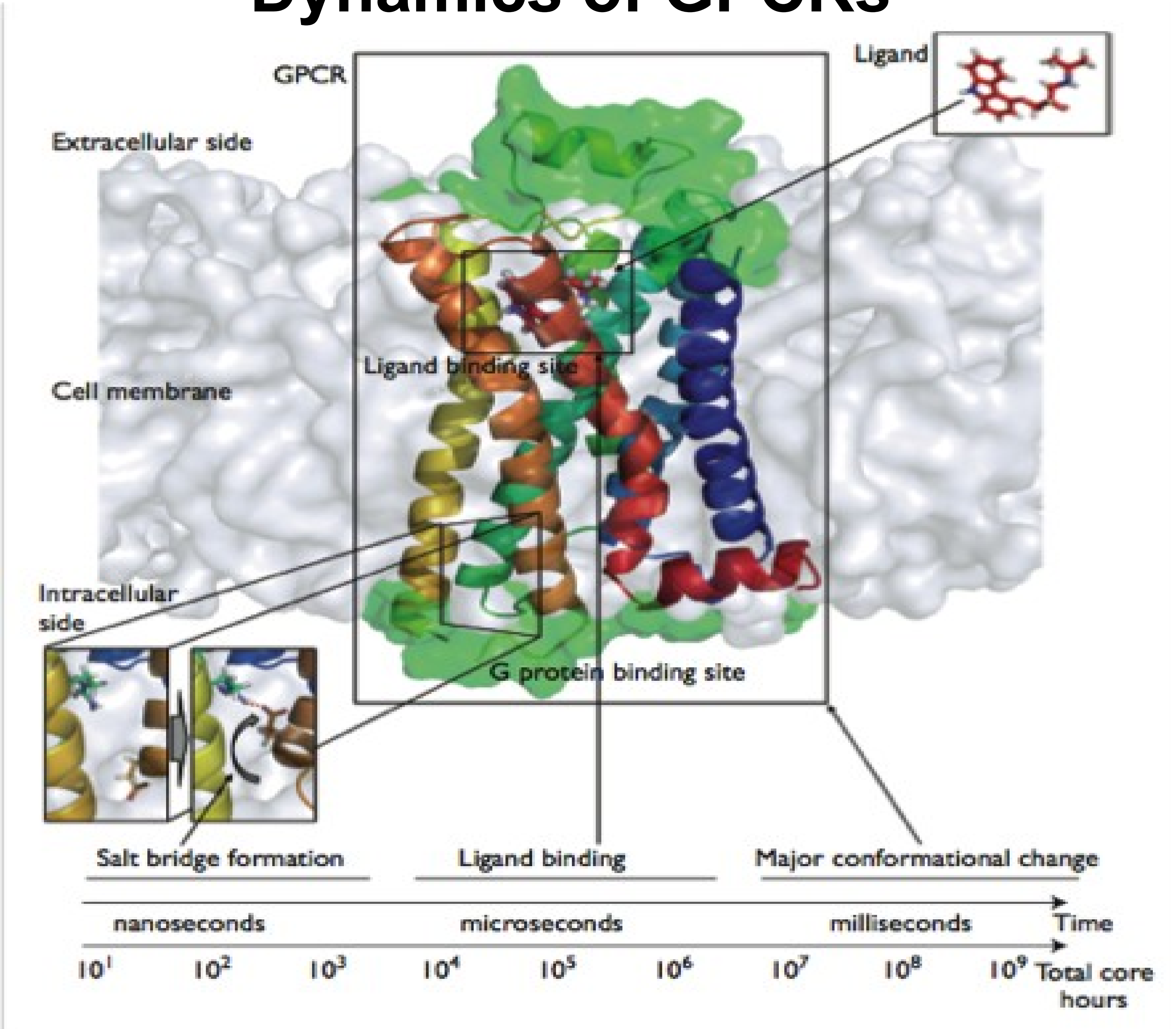
We're pleased to announce a new academic research grant program: Google Exacycle for Visiting Faculty. Through this program, we'll award up to 10 qualified researchers with at least 100 million core-hours each, for a total of 1 billion core-hours. The program is focused on large-scale, CPU-bound batch computations in research areas such as biomedicine, energy, finance, entertainment, and agriculture, amongst others. For example, projects developing large-scale genomic search and alignment, massively scaled Monte Carlo simulations, and sky survey image analysis could be an ideal fit.

G-Protein Coupled Receptors (GPCR)

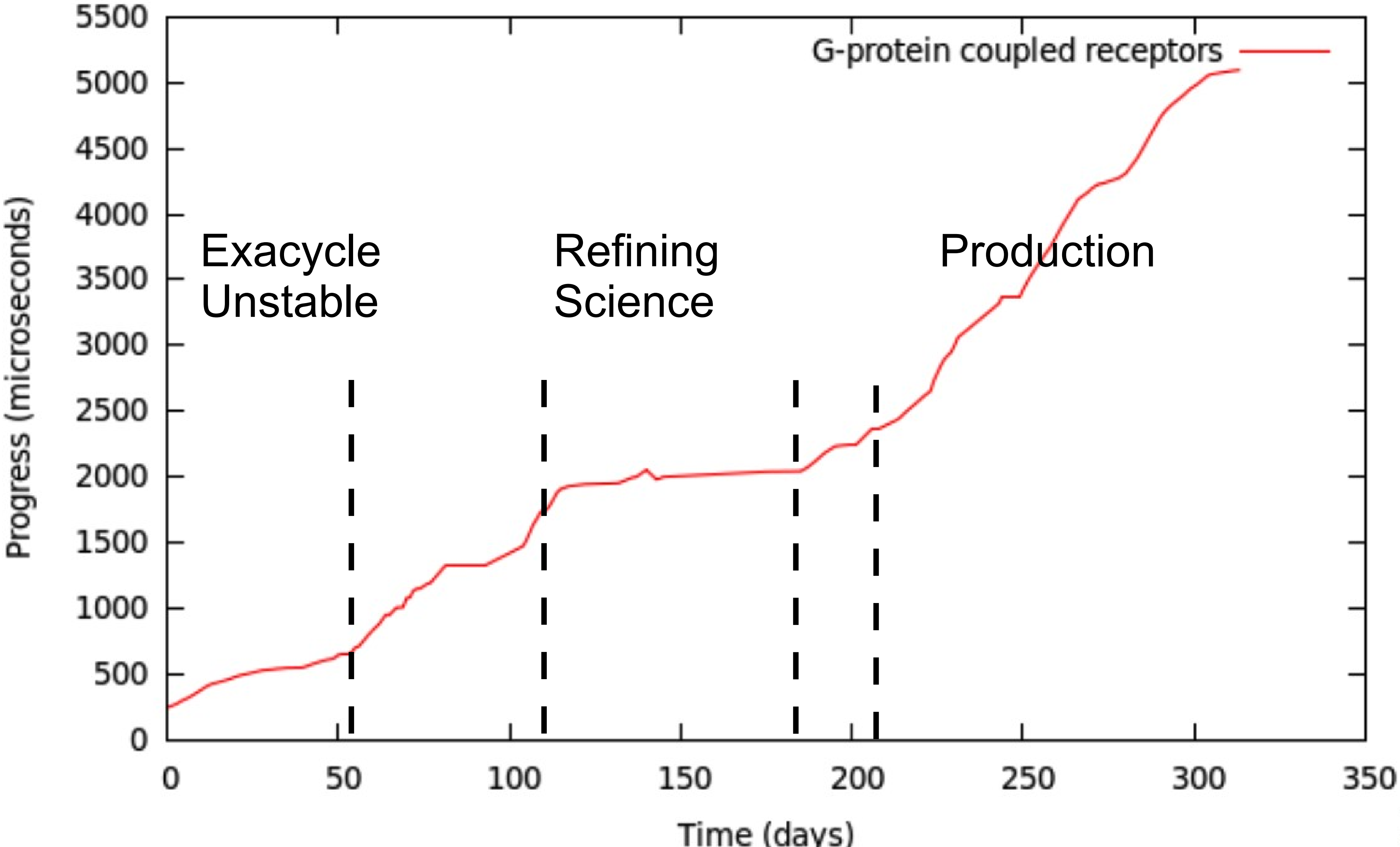


- GPCRs enable the exchange of molecules between cells and their environment.
- 40% of pharmaceuticals target GPCRs.

Dynamics of GPCRs

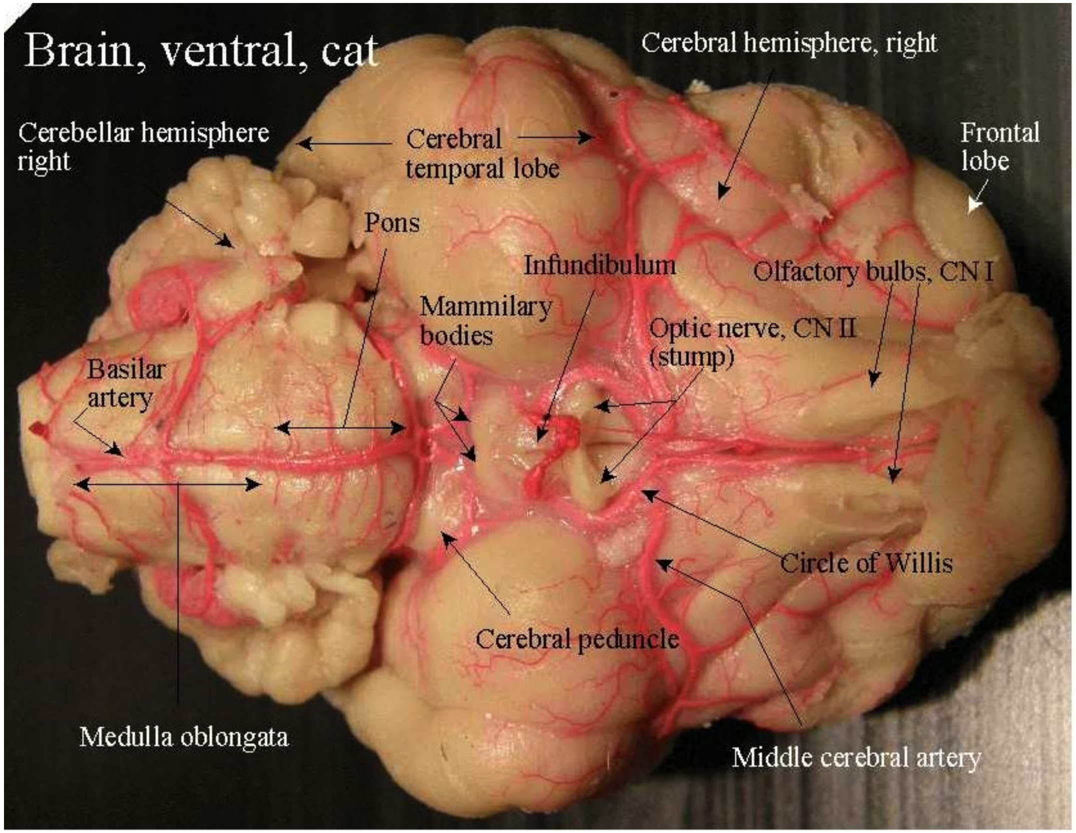


Experience with GPCR Simulations

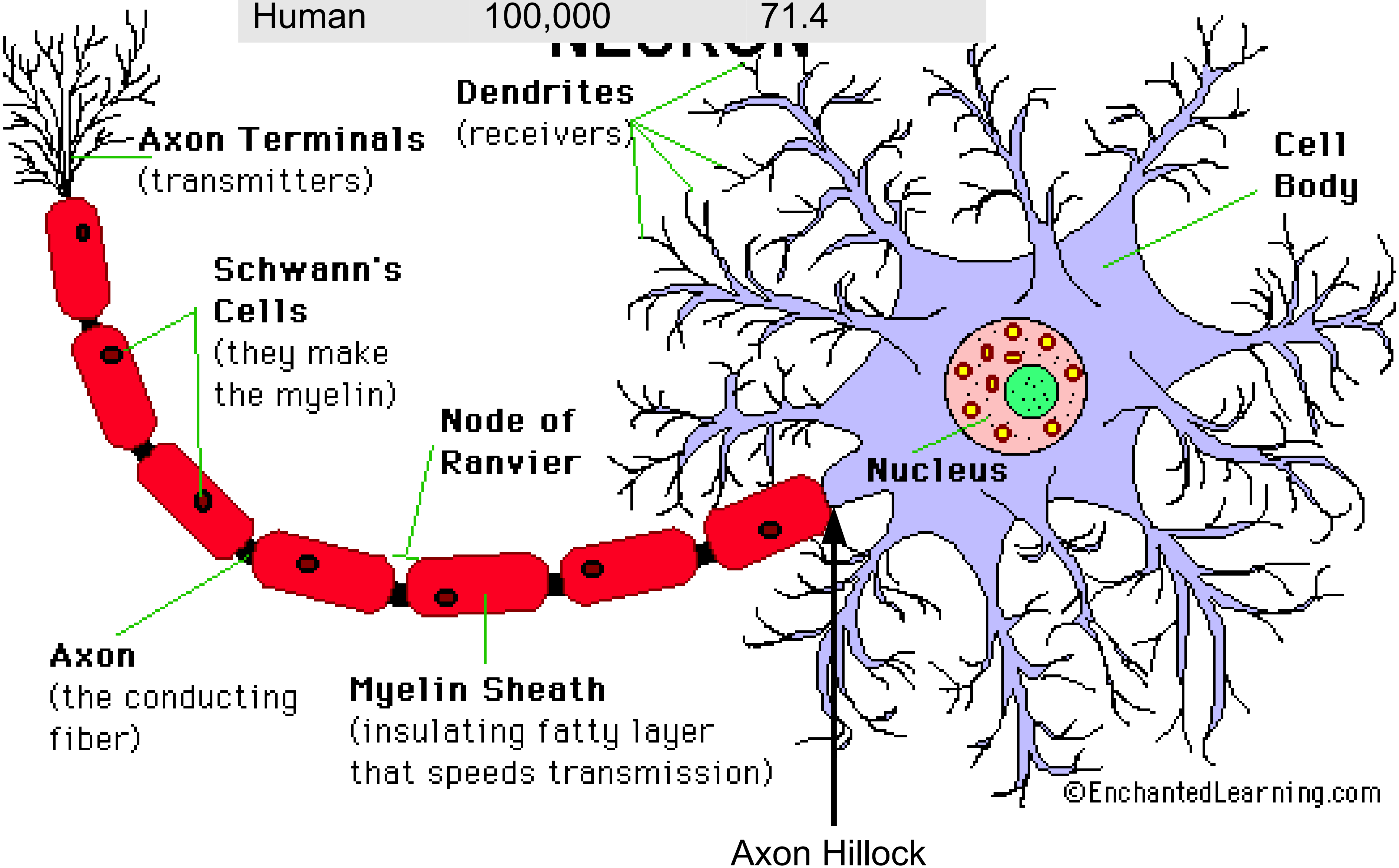


Brain Simulation Case Study

- Brain science 101
- Simulating sleep wake cycles in the cat brain.



Animal	Neurons(M)	Neuron-M per brain-g
Sea slug	0.01	
Mouse	4	10
Cat	300	10
Human	100,000	71.4



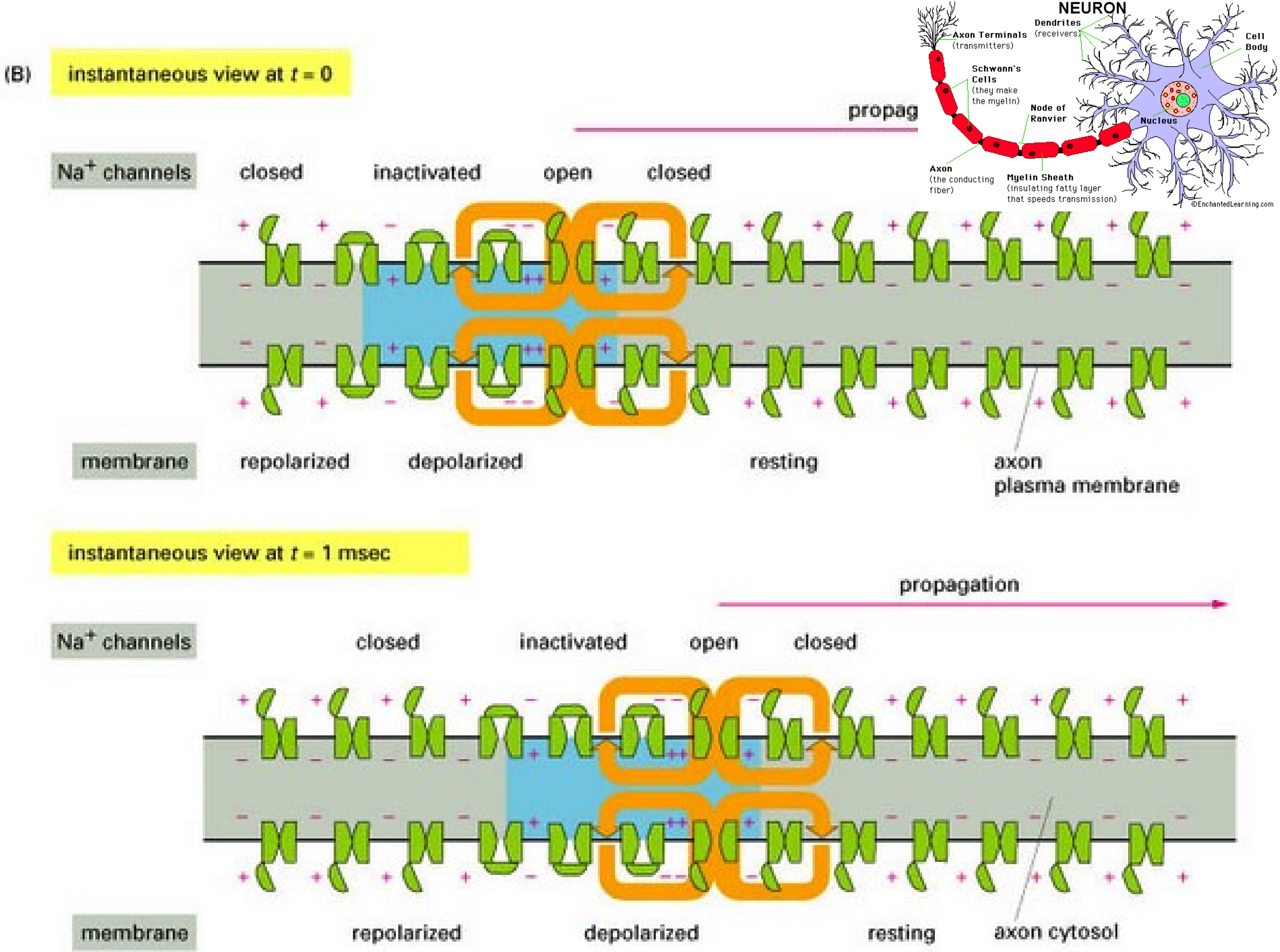
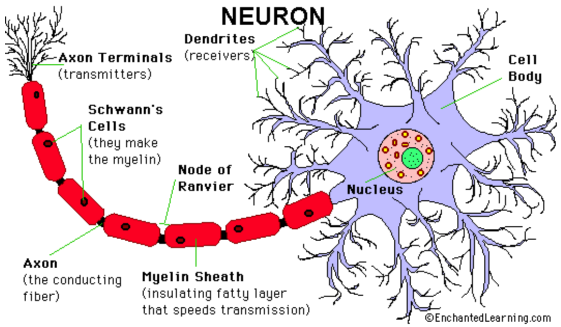
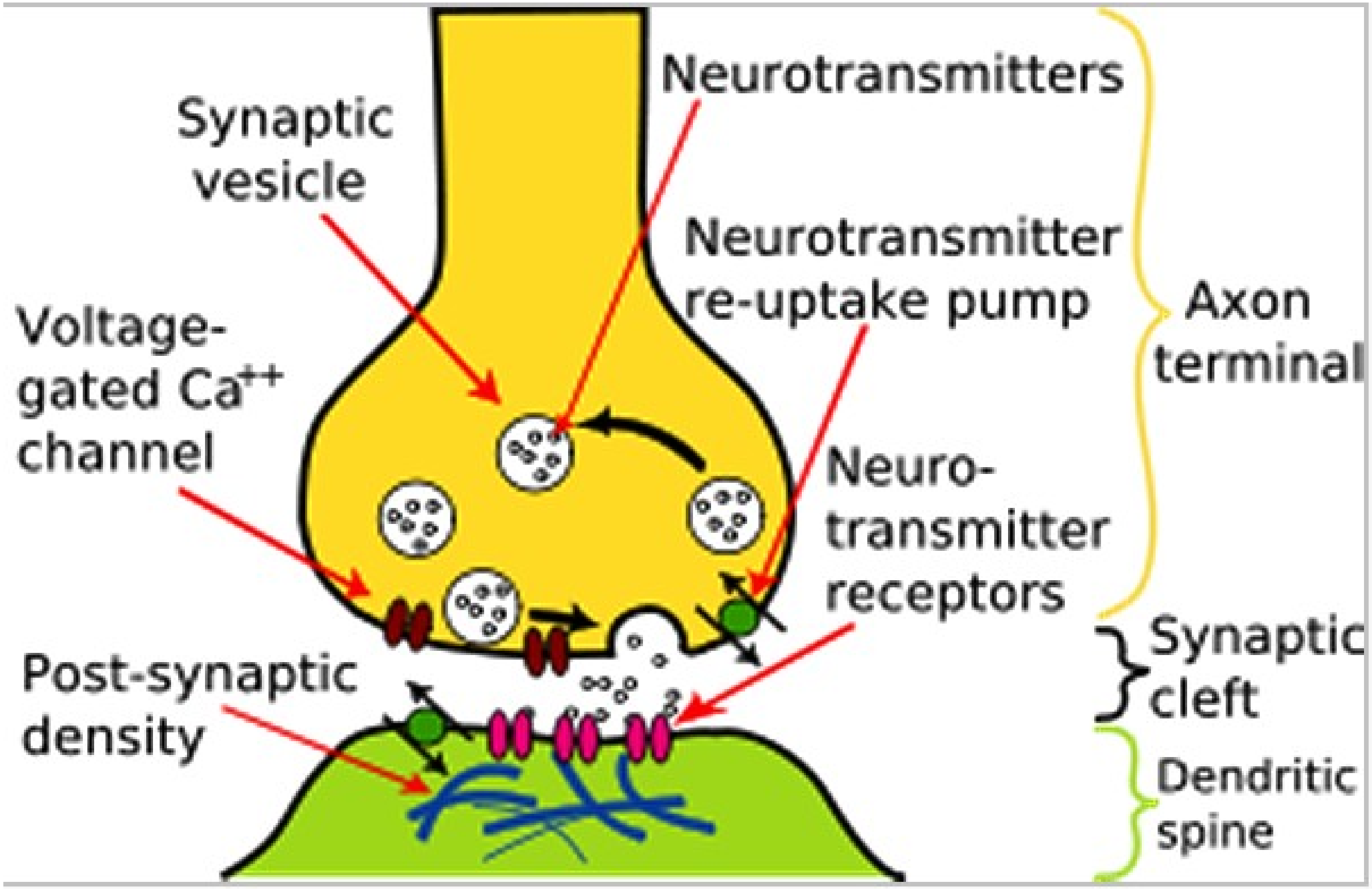


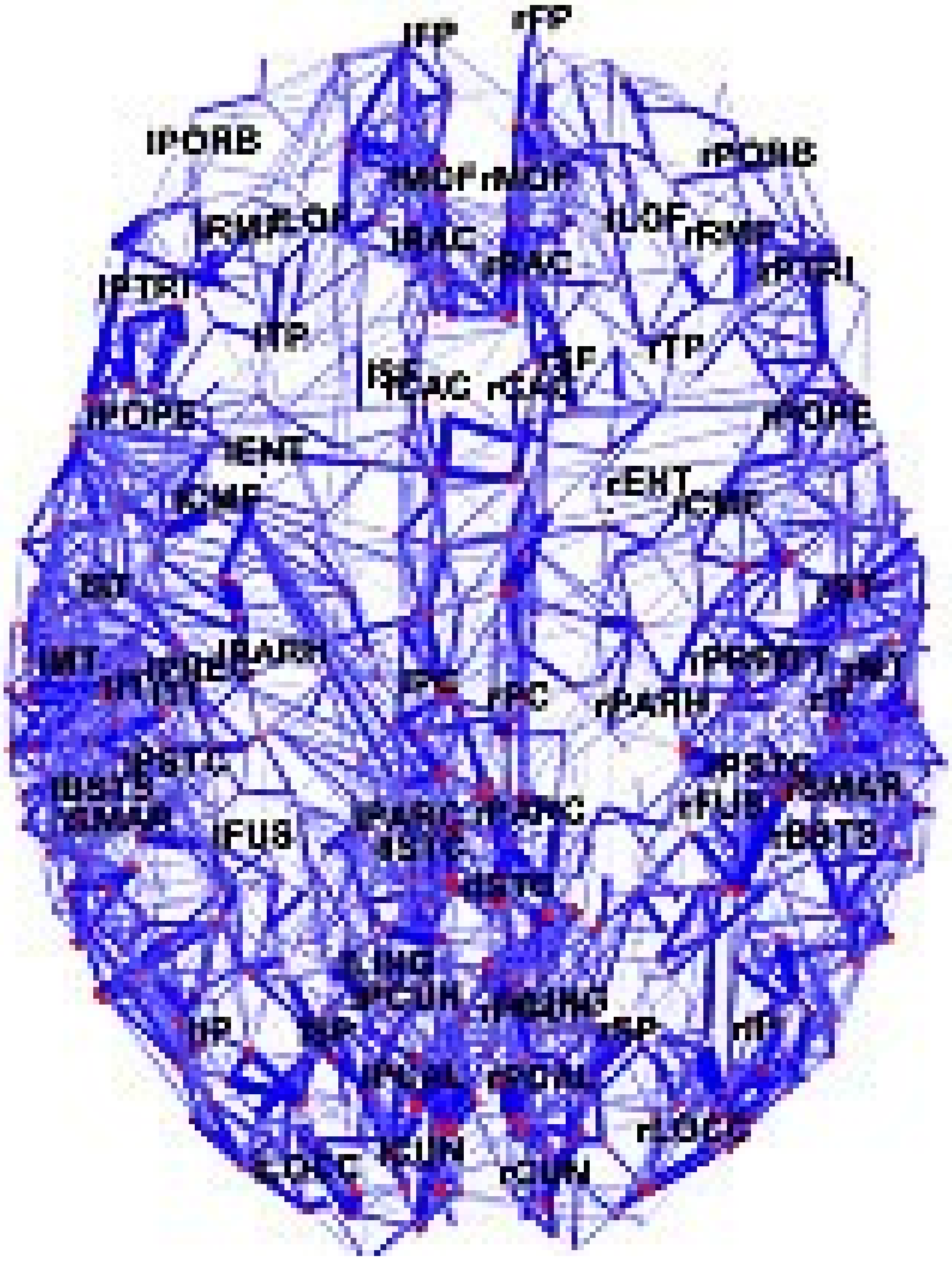
Figure 12-38 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Synapses Connect Neurons
#synapses ~ 1,000 X #Neurons

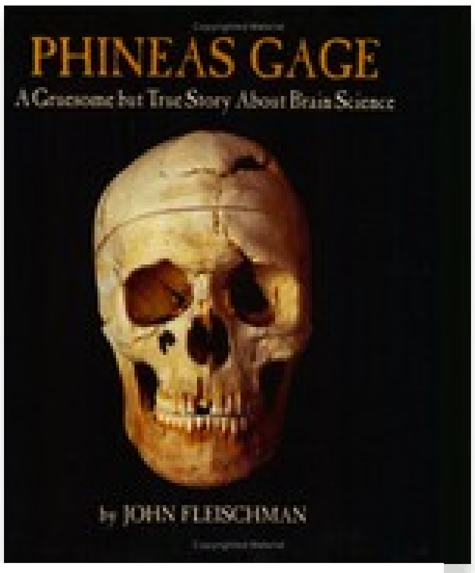




Examples of High Level Brain Wirings



Mapping Function to Brain Anatomy

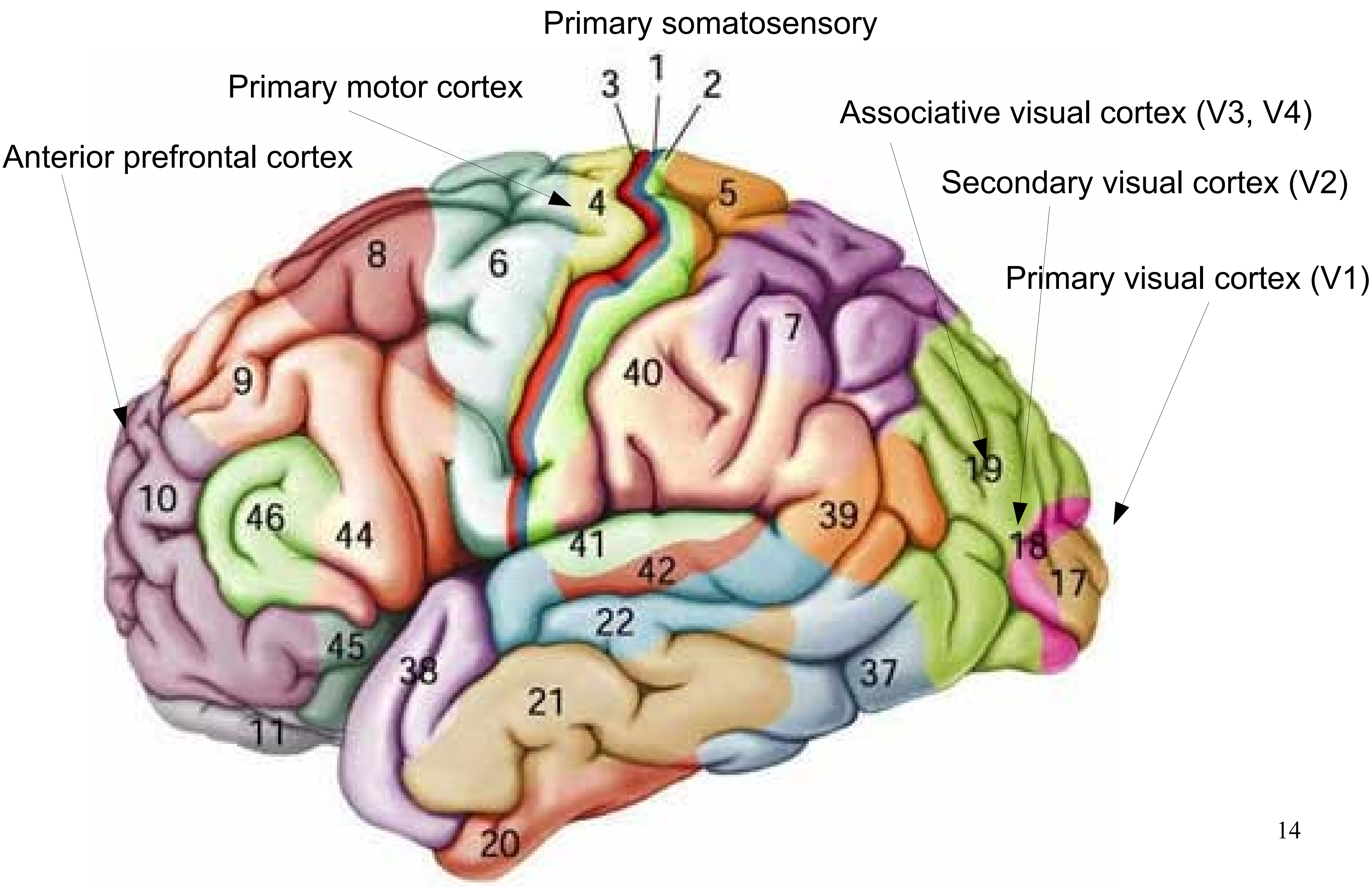


Phineas Gage: A Gruesome but True Story About Brain Science

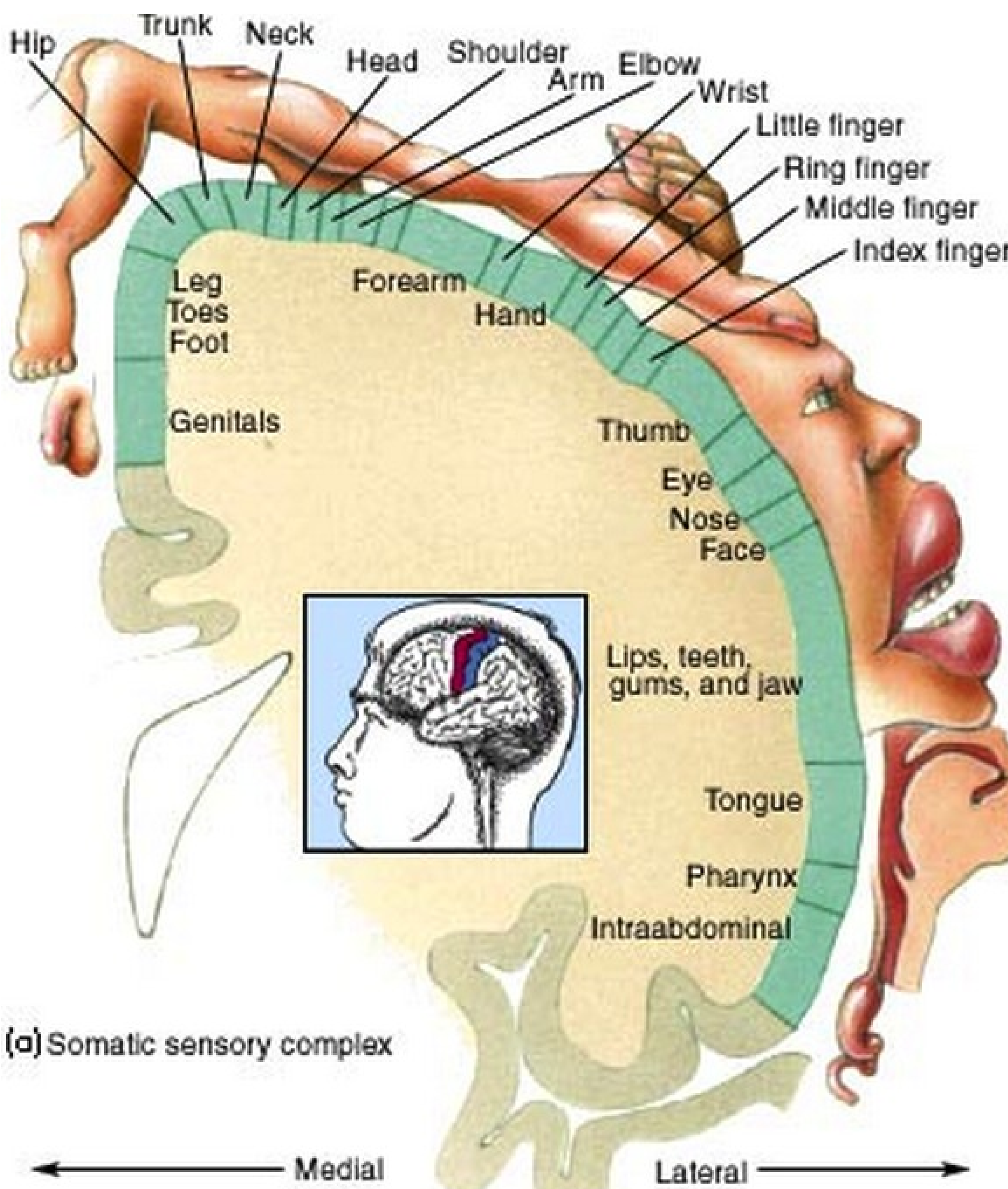
by John Fleischman

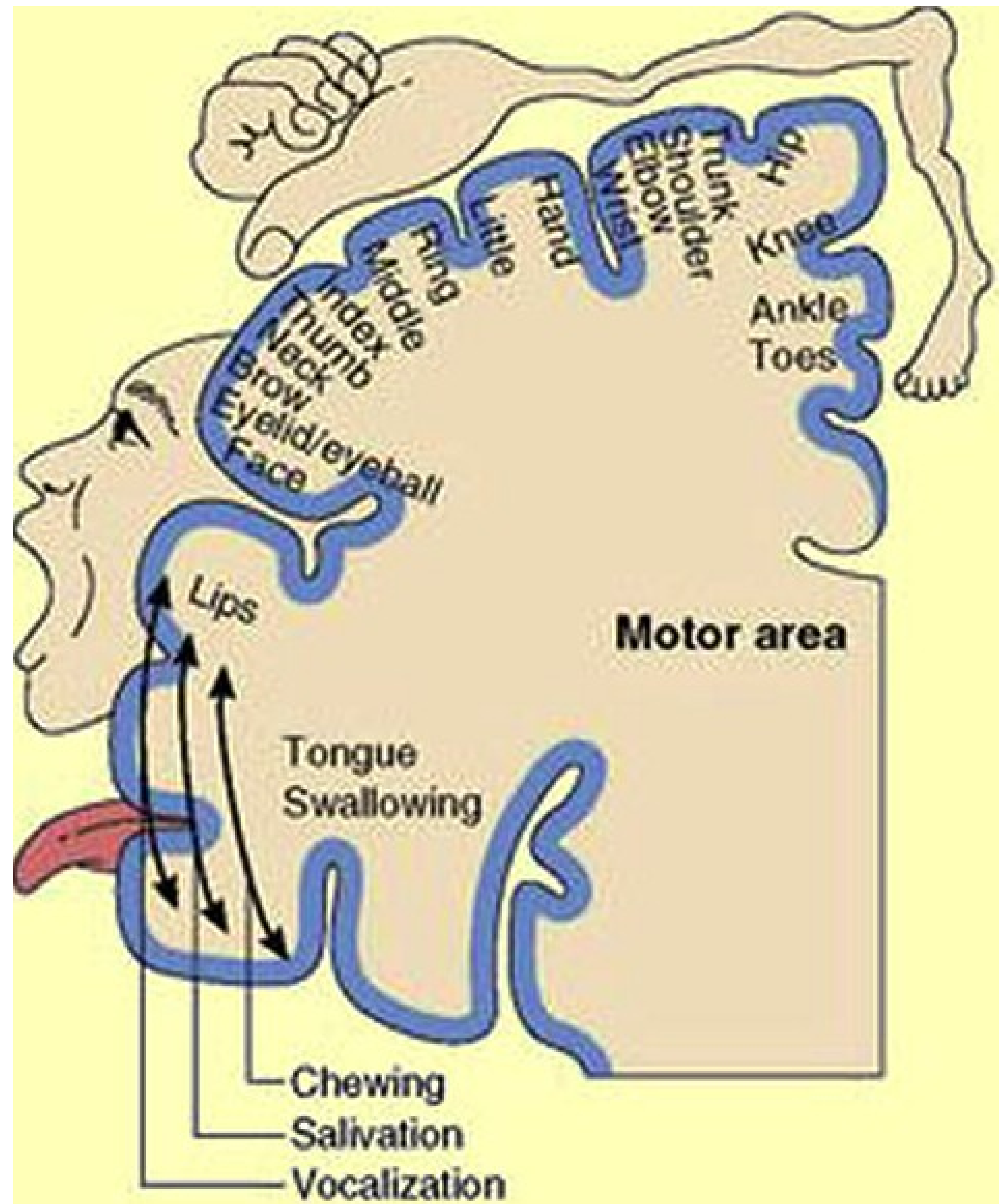
Phineas Gage was truly a man with a hole in his head. A railroad construction foreman, Phineas was blasting rock near Cavendish, Vermont, in 1848 when a thirteen-pound iron rod was shot through his brain. Miraculously, he survived another eleven years and became a textbook case in brain science. But he was forever changed by the accident, and what happened inside his brain...more

Brodmann's Areas in the Human Brain



Details of Sensory Motor Areas





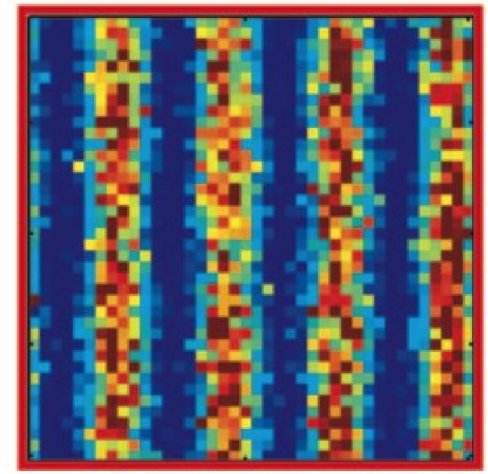
J Neurophysiol 93: 1671–1698, 2005.
First published November 10, 2004; doi:10.1152/jn.00915.2004.

Modeling Sleep and Wakefulness in the Thalamocortical Sy

Sean Hill and Giulio Tononi
Department of Psychiatry, University of Wisconsin—Madison, Madison, Wisconsin

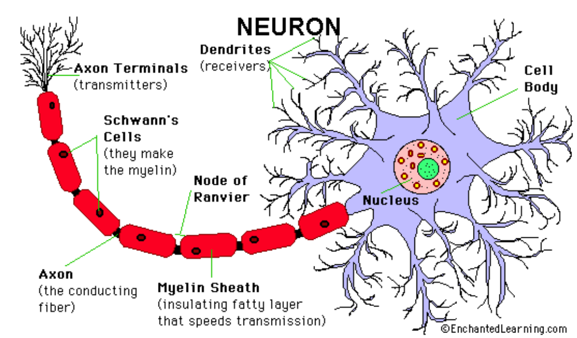
Submitted 1 September 2004; accepted in final form 26 October 2004

Vertical



- Can a brain simulation reproduce sleep-wake cycles?

- Model
  - Point model of 65K neurons
  - Statistical model of neural circuits
  - Calibrate from empirical data



Model Parameterizations

- Neurons (65K)
  - Ion channels, propagation delays, types
- Synapses (~65M)
  - Neurotransmitter, weights on neurons, connections

Hundreds of parameters!!!

Models of Neuron Firings

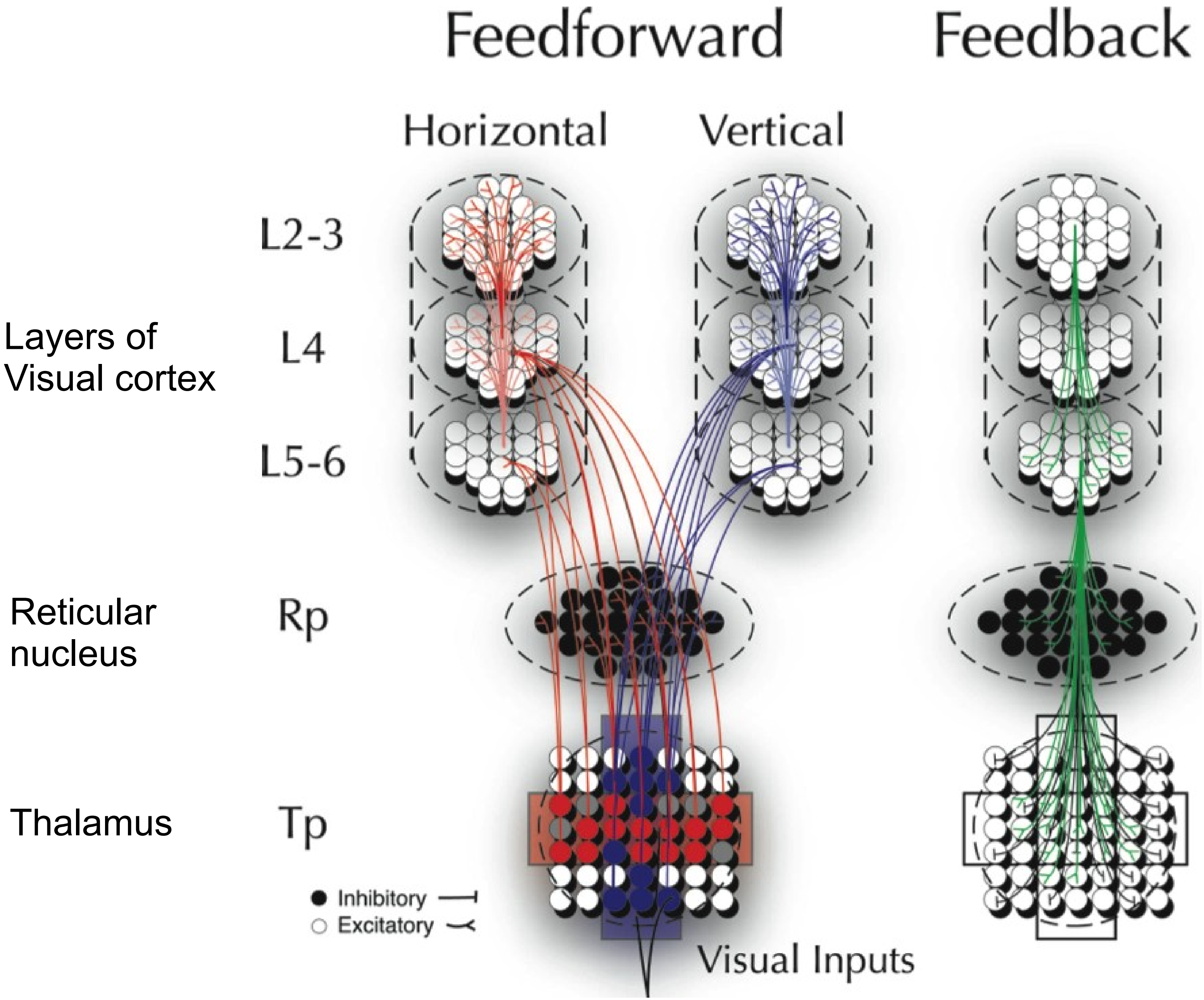
A dynamic threshold (θ) is defined for each cell that determines at which membrane potential the cell should fire

$$\frac{d\theta}{dt} = -(\theta - \theta_{eq})/\tau_\theta$$

The change in membrane potential V for each neuron is as follows

$$\frac{dV}{dt} = [-g_{NaL}(V - E_{Na}) - g_{KL}(V - E_K) - I_{syn} - I_{int}]/\tau_m - g_{spike}(V - E_K)/\tau_{spike}$$

where the conductances for the sodium leak ($g_{NaL} = 0.2$) and potassium leak ($g_{KL} = 1.0$–1.85) are the primary determinants of the resting membrane potential. Conductance units are dimensionless


Feedforward
Feedback
Horizontal
Vertical
Layers of
Visual cortex
L2-3
L4
L5-6
Reticular
nucleus
Rp
Thalamus
Tp
Inhibitory
Excitatory
Visual Inputs

Study Summary

- Discovered simulation parameters that: (a) are consistent with empirical sleep-wake cycles and (b) satisfy a set of biological constraints.
- But...
  - Is the model over-fitted?
  - Very time-consuming to explore the parameter space to determine if there is a consistent parameter set.
  - Ad hoc methodology
    - How systematically build model?
    - How diagnose model errors (e.g., concept of diagnostics in statistics)?
  - Little engineering rigor
    - No concept of tests that relate simulation results to biology

Why Science in the Cloud?

- Burst capacity
  - Access to many thousands of cores
- Reproducibility
  - Investigators use the same computational tools and data
- Sharing
  - Build on the results of others
- Efficient use of scarce research dollars
  - Avoid investments in infrastructure with a short lifetime

System Challenges

- Fine grain parallelism on a commodity infrastructure
- Low friction scaling
  - Scientists should focus on science not programming distributed systems.
- Interactive exploratory analysis at scale.
- Introspective batch processing.
- Multi-cloud support
  - Data are big & distributed
  - Commercial science requires both public and private` clouds

Science Methodology Challenges

- Systematic model development that is calibrated with empirical data
- Testing methodology
- Integrate models at multiple levels of granularity
- Standardized schemas so can share data
  - But accommodate rapid evolution of knowledge
- Create a culture of software engineering among scientists
  - Software is the lab notebook for computational discovery