International Journal of Neural Systems, Vol. 16, No. 3 (2006) 1–12

World Scientific
www.worldscientific.com

COMPUTATIONAL NEUROGENETIC MODELLING: A PATHWAY TO NEW DISCOVERIES IN GENETIC NEUROSCIENCE

LUBICA BENUSKOVA*, VISHAL JAIN†, SIMEI G. WYSOSKI‡
and NIKOLA K. KASABOV§
Knowledge Engineering & Discovery Research Institute,
Auckland University of Technology, Ronald Trotter Building,
581-585 Great South Road, Penrose, Auckland, New Zealand
**lbenusko@aut.ac.nz*
†*vjain@aut.ac.nz*
‡*swysoski@aut.ac.nz*
§*nkasabov@aut.ac.nz*

The paper presents a methodology for using computational neurogenetic modelling (CNGM) to bring new original insights into how genes influence the dynamics of brain neural networks. CNGM is a novel computational approach to brain neural network modelling that integrates dynamic gene networks with artificial neural network model (ANN). Interaction of genes in neurons affects the dynamics of the whole ANN model through neuronal parameters, which are no longer constant but change as a function of gene expression. Through optimization of interactions within the internal gene regulatory network (GRN), initial gene/protein expression values and ANN parameters, particular target states of the neural network behaviour can be achieved, and statistics about gene interactions can be extracted. In such a way, we have obtained an abstract GRN that contains predictions about particular gene interactions in neurons for subunit genes of AMPA, $GABA_A$ and NMDA neuro-receptors. The extent of sequence conservation for 20 subunit proteins of all these receptors was analysed using standard bioinformatics multiple alignment procedures. We have observed abundance of conserved residues but the most interesting observation has been the consistent conservation of phenylalanine (F at position 269) and leucine (L at position 353) in all 20 proteins with no mutations. We hypothesise that these regions can be the basis for mutual interactions. Existing knowledge on evolutionary linkage of their protein families and analysis at molecular level indicate that the expression of these individual subunits should be coordinated, which provides the biological justification for our optimized GRN.

Keywords: Computational neurogenetic modelling; protein sequence analysis; gene-brain relationship; pattern discovery; AMPAR; GABRA; NMDAR.

1. Introduction

A specific gene from the genome relates to the activity of a neuronal cell by means of a specific protein. Complex interactions between genes and proteins within the internal gene/protein regulatory network influence the functioning of each neuron and a neural network as a whole.[1] Even in the presence of a mutated gene in the genome, which is known to cause a brain disease, the neurons can still function normally provided a certain pattern of interaction between genes is maintained. For instance, a preliminary screen using complementary DNA arrays and brain tissue from pre-seizure genetic absence epilepsy rats from Strasbourg (GAERS) has failed to identify significant changes in gene expression compared to control samples.[2] On the other hand we can assume, if there are no mutated genes, abnormalities in the brain functioning can be observed as a consequence of abnormal interaction between genes. With the advancement of molecular research technologies huge amount of data and information are available about the genetic basis of neuronal functions and diseases.[1,3] This information can be utilized to create models of brain functions and diseases that include models of gene interactions within models of neural

networks. We call this new area *computational neurogenetic modelling* (CNGM), the principles of which have been introduced in Refs. 4–6.

To understand human brain function we need to integrate knowledge from genomics, proteomics, neuroscience, psychology, and theoretical disciplines such as computer science and physics. To understand the function of genes, we need to know how they are expressed, when they are expressed, conservation among their products and their response to therapeutic drugs. In addition, we also need to know how they interact and influence the dynamics of neural networks. For this purpose, we develop and study CNGM. In general, we can say that CNGM potentially may help to understand a broad range of fundamental questions about genetic influences on mental processes and diseases. The use of such models can facilitate new discoveries in the area of genetic neuroscience, which is illustrated in this paper.

Genetic neuroscience uses principles from cellular and molecular biology to investigate questions about gene actions in neurons. Prominent approaches in this area are genetic mutations, gene knockouts, protein purification, gene expression analysis, GAS chromatography, gel electrophoresis, high performance liquid chromatography (HPLC), western blotting, etc. Another multigene approach, high throughput genotyping technology, can be used to define single nucleotide polymorphisms that characterize diseases. However, all these techniques are useful but time consuming and are very expensive. Thus, for providing biological insight into our CNGM, we use standard bioinformatics operating strategies like studying neuronal protein sequence conservation/variation. Vast comparative analysis of crucial protein sequences and structures can also reveal evolutionary relationships between specific brain regions of humans and other species, increasing the relevance of CNGM.

In this paper we present a novel methodology of using CNGM for modelling gene influences upon the electrical activity of the brain, analyse critical genes and gene interaction patterns and finally integrate extracted knowledge from CNGM simulations with currently known biological facts.

2. Formulation of an Abstract CNGM

Proteins that mediate general cellular or specific information-processing functions in neurons are usually complex molecules comprised of several subunits, each of them being coded by a separate gene.[7] Based on our protein sequence analysis (see below) and considering biological knowledge on neuro-receptors,[8,9] we consider one protein P_j to be one functional unit coded for by one group of related genes, G_j. In general, we consider two sets of gene groups — a set $\mathbf{G}_{\text{gen}}$ that relates to general cell proteins and a set $\mathbf{G}_{\text{spec}}$ that defines specific neuronal information-processing proteins (e.g. receptors, ion channels, etc.). The two sets form together a set $\mathbf{G} = \{G_1, G_2, \ldots, G_n\}$. We assume that the expression level of each gene group $g_j(t+\Delta t)$ (being the sum of expression levels of all the genes in that group) is a nonlinear function of expression levels of all the gene groups in $\mathbf{G}$. This dependence is inspired by discrete models from Refs. 10 and 11, with the discrete delay Δt representing regular time intervals at which gene expression data would be gathered in the real experiment, i.e.,

$$g_j(t+\Delta t) = \sigma\left(\sum_{k=1}^{n} w_{jk} g_k(t)\right) \tag{1}$$

We work with normalized gene expression values, expressed by a nonlinear sigmoid $\sigma(\mathrm{x}) \in (0,1)$. The coefficients $w_{ij} \in (-5,5)$ are elements of the square matrix $\mathbf{W} = \{w_{ij}\}$ of gene interaction weights. The latter borders have been chosen experimentally in order to lead to various types of nonlinear dynamics, i.e. constant as well as periodic, quasi-periodic and chaotic. It is important to identify what the sign of an interaction means, i.e. whether $w_{ij} > 0$ or $w_{ij} < 0$. The positive interaction means that upregulation (downregulation) of one gene group supports the upregulation (downregulation) of the other gene group, respectively. The negative interaction means that upregulation (downregulation) of one gene group leads to the downregulation (upregulation) of the other gene group, respectively. Initial values of gene expressions are small random values, i.e., $g_j(0) \in (0, 0.1)$. We assume a linear relationship between protein levels and gene expression levels, thus the protein level $p_j(t+\Delta t)$ reads

$$p_j(t+\Delta t) = \left(p_j^{\max} - p_j^{\min}\right)\sigma\left(\sum_{k=1}^{n} w_{jk} g_k(t)\right) + p_j^{\min} \tag{2}$$

where $p_j^{\min}$ and $p_j^{\max}$ are minimal and maximal levels of protein p_j, respectively. The delay Δt again corresponds to the interval of protein expression data gathering. These delays are the same as for the Eq. (1). This is based on the fact that protein expression data are being gathered for all proteins of interest at the same time points as the gene expression data. It is a common practice to derive **W** based on all gene expression data being collected at the same time intervals Δt.[12] The linear relationship in Eq. (2) is based on findings that protein complexes, which have clearly defined interactions between their subunits, have highly correlated levels with mRNA expression levels.[13,14] Subunits of the same protein complex show significant co-expression, both in terms of similarities of absolute mRNA levels and expression profiles, e.g., subunits of a complex have correlated patterns of expression over a time course.[13] This implies that there should be a correlation between mRNA and protein concentration, as these subunits have to be available in stoichiometric amounts for the complexes to function.[14] And this is exactly the case of proteins in our model, which are receptors and ion channels, comprised of respective ratios of subunits. Thus, the use of Eq. (2) is justified in our model. In our model, some protein levels are directly related to the values of neuronal parameters P_j such that

$$P_j(t) = P_j(0)p_j(t) \tag{3}$$

where $P_j(0)$ is the initial value of the neuronal parameter at time $t = 0$, and $p_j(t)$ is the corresponding protein level at time t. In such a way the gene/protein dynamics is linked to the dynamics of ANN. Neuronal parameters and their corresponding proteins are summarized in Table 1.

The choice of neural parameters depends on a task, which we want to simulate and knowledge from molecular neurobiology. In our case, let the simulated neural behaviour be the brain electrical activity measured as EEG. Support for this approach comes from recent studies that have shown that brain electrical oscillations are genetically determined and differ between individuals and families.[15–17] Spectral characteristics of electrical oscillations are highly conserved in one individual but vary between individuals even when recorded under the same conditions. Genetically related individuals have similar features in electrical activity of their brains.

Table 1. Our CNGM neuron parameters and related proteins.

Neuron's parameter	Protein
Amplitude and time constants of:	
Fast excitation	AMPAR
Slow excitation	NMDAR
Fast inhibition	GABRA
Slow inhibition	GABRB
Firing threshold	SCN, KCN, CLC

Abbreviations: AMPAR = (amino-methylisoxazole-propionic acid) AMPA receptor, NMDAR = (*N*-methyl-D-aspartate acid) NMDA receptor, GABRA = (gamma-aminobutyric acid) $GABA_A$ receptor, GABRB = $GABA_B$ receptor, SCN = Sodium voltage-gated channel, KCN = kalium (potassium) voltage-gated channel, CLC = chloride channel. In the case of firing threshold, Eq. (3) is modified to include three proteins (ion channels) SCN, KCN and CLC.

However, genes activate and produce protein molecules over hours, whereas brain local field potentials and thus EEG change over milliseconds. The question arises how we will integrate this slow gene/protein dynamics with the fast electric dynamics of the model neural network. Our target signal will be the resting human EEG. Spectral characteristics of resting EEG are preserved irrespectively of what time of aroused resting wakefulness during the day the signal is recorded. Thus any point of daily gene dynamics should lead to the resting electrical signal with the same spectral characteristics. Therefore we will compress the daily gene dynamics into a time scale of the EEG measurement interval, that is T = 1 min, in order to optimize the desired gene interaction matrix **W**. As a result of the compression of gene dynamics into the time interval of LFP/EEG measurement, we set up the simulated interval of gene and protein changes Δt to 1 s, whereas the time step of updating of neural network dynamics will be 1 ms.

Besides the genes coding for the proteins listed in Table 1 (that directly influence the neuronal spiking activity), we have included in our GRN nine more genes that are not directly linked to neuronal information-processing parameters. These genes are: c-jun, mGLuR3, Jerky, BDNF, FGF-2, IGF-I, GALR1, NOS, S100beta. These other genes are assumed to interact with genes and proteins listed in Table 1, since their expression depends

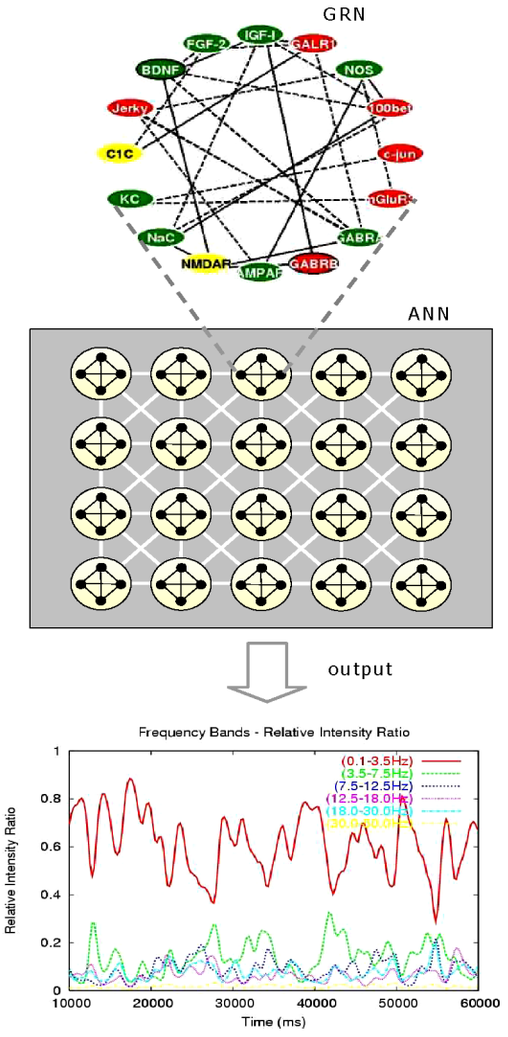


Fig. 1. Computational NeuroGenetic Model (CNGM) as an Artificial Neural Network (ANN) with particular output behaviour.

on neural activity.[18–24] We have included them for future studies. For now, they only represent the generality of our approach. CNGM has been illustrated as an ANN with particular output behaviour (see Fig. 1).

The CNGM from Eqs. (1)–(3) is a general one and can be integrated with the spiking neural network model or any other neural network model. In the current form we have made these simplifying assumptions:

(i) Each neuron has the same gene regulatory network (GRN), i.e., the same genes and the same interaction gene matrix $\mathbf{W}$, and starts from the same initial values of gene expressions.
(ii) Feedbacks from neuronal activity or any other external factors to gene expression levels are not explicitly considered. They can be incorporated in Eq. (1) by means of a special term.

To simulate brain electrical activity, in a particular implementation of the abstract CNGM, we have used the spiking neural network model, which is discussed in a later section.

3. Molecular Sequence Analysis of Excitatory and Inhibitory Receptors: Motivation for the Creation of CNGM

We have investigated through sequence analysis, the excitatory and fast inhibitory receptors (i.e., AMPAR, NMDAR and GABRA) gene groups, which were found participating in statistically significant gene interactions in the presented CNGM of LFP/EEG generation. We mainly focused on them because they are identified as crucial in CNGM simulations and also they play a major role in most of the mental disorders through their direct or indirect interactions with several other genes/proteins and through specific parameter functions (e.g., excitation and inhibition). Each of these proteins is comprised of several subunits, and each subunit is coded by a separate gene (see Table 2).

Initially, the information related to expression of these subunit genes, mutations, etc. was collected through relevant literature survey followed by sequences retrieval from NCBI database. We did preliminary analysis for searching the common motifs (patterns that occurs repeatedly in a group of related protein or DNA sequences) among these subunits. Through this initial analysis we observed that all detected motifs were belonging to similar protein families, like ligand-gated ion channel, receptor family ligand binding region and neurotransmitter-gated ion channel ligand binding domain. This observation motivated us for performing the detailed residual inspection. For our investigation, we employed the standard operating bioinformatics procedure, i.e., comparative analysis from similarity measures that is widely accepted approach by most of neuroscientists. The method is familiarly known as multiple sequence alignments (MSA). Scope of

Table 2. List of subunit proteins for AMPA, GABRA and NMDA receptors that define specific neuronal information-processing parameter functions.

gi|number|	Human protein	Gene	Sequence length (aa)
gi|1169959|	Glutamate receptor ionotropic, AMPA 1; (GluR-1) (GluR-A) (GluR-K1)	GRIA1	906
gi|23831146|	Glutamate receptor ionotropic, AMPA 2; (GluR-2) (GluR-B) (GluR-K2)	GRIA2	883
gi|1169961|	Glutamate receptor ionotropic, AMPA 3; (GluR-3) (GluR-C) (GluR-K3)	GRIA3	894
gi|1346142|	Glutamate receptor ionotropic, AMPA 4; (GluR-4) (GluR4) (GluR-D)	GRIA4	902
gi|11496971|	NMDA receptor 1 isoform NR1-1 precursor	GRIN1	885
gi|14285603|	Glutamate [NMDA] receptor subunit epsilon 1 precursor; (NR2A) (NMDAR2A)	GRIN2	1464
gi|14548162|	Glutamate [NMDA] receptor subunit epsilon 2 precursor; (NR2B) (NMDAR2B)	GRIN3	1484
gi|2492629|	Glutamate [NMDA] receptor subunit epsilon 3 precursor; (NR2C) (NMDAR2C)	GRIN4	1233
gi|18201966|	Glutamate [NMDA] receptor subunit epsilon 4 precursor; (NR2D) (NMDAR2D)	GRIN5	1336
gi|38327554|	Gamma-aminobutyric acid (GABA) A receptor, alpha 1 precursor	GABRA1	456
gi|1346078|	Gamma-aminobutyric-acid receptor alpha-2 subunit precursor	GABRA2	451
gi|4557603|	Gamma-aminobutyric acid A receptor, alpha 3 precursor	GABRA3	492
gi|1346079|	Gamma-aminobutyric-acid receptor alpha-4 subunit precursor	GABRA4	554
gi|399519|	Gamma-aminobutyric-acid receptor alpha-5 subunit precursor	GABRA5	462
gi|23831128|	Gamma-aminobutyric-acid receptor beta-1 subunit precursor	GABRB1	474
gi|455946|	Gamma-aminobutyric acid A receptor beta 2 subunit	GABRB2	474
gi|120773|	Gamma-aminobutyric-acid receptor beta-3 subunit precursor	GABRB3	473
gi|27820121|	Gamma-aminobutyric-acid receptor gamma-1 subunit precursor	GABRG1	465
gi|38788155|	Gamma-aminobutyric acid A receptor, gamma 2 isoform 1 precursor	GABRG2	475
gi|13959689|	Gamma-aminobutyric-acid receptor gamma-3 subunit precursor	GABRG3	467

		Position			
Gene	...	**260 270**	...	**350 360**	...
GABRA3	...	IGEYVVMIIHFHLKRKIGYFV	...	AVCYAFVFSALIEFAIVNYFI	...
GABRA5	...	IGEYIIMIAHFHLKRKIGYFV	...	AVCYAFWFSALIEFAIVNYFI	...
GABRA1	...	IGEYVVMIIHFHLKRKIGYFV	...	AVCYAFVFSALIEFAIVNYFI	...
GABRA2	...	IGEYIVMIAHFHLKRKIGYFV	...	AVCYAFVFSALIEFAIVNYFI	...
GABRA4	...	IGEYIVMIVYFHLRRKMGYFM	...	AVCFAFVFSALIEFAAVNYFI	...
GABRG1	...	SGDYVIMIIFFDLSRRMGYFI	...	SVCFIFVFAALMEYGILHYFI	...
GABRG2	...	SGDYVVMSVYFDLSRRMGYFI	...	SVCFIFVFSALVEYGILHYFV	...
GABRG3	...	AGDYVVMIIYFELSRRMGYFI	...	IVCFLFVFAALMEYAILNYYS	...
GABRB2	...	IGSYPRLSLSFKLKRNIGYFI	...	MGCFVFVFMALLEYALVNYIF	...
GABRB3	...	IGAYPRLSLSFRLKRNIGYFI	...	MGCFVFVFLALLEYAFVNYIF	...
GABRB1	...	IGAYPRLSLSFRLKRNIGYFI	...	MGCFVFVFLALLEYAFVNYIF	...
GRIA2	...	PQKSKPGVFSFLDPLAYEIWM	...	LIIISSYIANLAAFLIVERMV	...
GRIA3	...	PQKSKPGVFSFLDPLAYEIWM	...	LIIISSYIANLAAFLIVERMV	...
GRIA4	...	PQKSKPGVFSFLDPLAYEIWM	...	LIIISSYIANLAAFLIVERMV	...
GRIA1	...	PQKSKPGVFSFLDPLAYEIWM	...	LIIISSYIANLAAFLIVERMV	...
GRIN2	...	.SNGIVSPSAFLEPFSASVWV	...	VIFLASYIANLAAFMIQEEFV	...
GRIN3	...	.SNGIVSPSAFLEPFSADVWV	...	VIFLASYIANLAAFMIQEEYV	...
GRIN4	...	.SNGIVSPSAFLEPYSPAVWV	...	VIFLASYIANLAAFMIQEQYI	...
GRIN5	...	.SNGIVSPSAFLEPYSPAVWV	...	VIFLASYIANLAAFMIQEEYV	...
GRIN1	...	EIPRS.ILDSFMQPFQSILWL	...	MIIVASYIANLAAFLVLDRPE	...

Fig. 2. Multiple alignments of all 20 subunits of 3 neuronal information-processing proteins showing consistent conservation of phenylalanine (F) at position 269 and leucine (L) at 353.

this bioinformatics technique is twofold: (a) gains understanding to identify the shared regions of homology; (b) determines the consensus sequence of several aligned sequences. For such strategies, there are however many softwares available free to academic users, but here in our case we used the CLUSTALW package (http://align.genome.jp/) developed by Koichi Ohkubo, (Genome Net), for details see Chenna.[25] A specific reason for picking this software was that it is mostly cited and secondly its output is a much readable representation than others in the field. Sequence similarity between all subunits was visualized, then we later used Box-Shade program V3.21 (http://

www.ch.embnet.org/software/BOX_form.html) in order to format our multiple alignment results for explanation purpose. Figure 2 is only a highlighted portion of most important observation and in general it indicates that comparison of the 20 sequences of subunit neuronal information-processing proteins has revealed a number of conserved residues. As a biological fact we know the number of structurally conserved residues increases with the binding site contact size, thus we expect similar function for these amino acids in this case also. Here we have observed the extent to which these conserved residues are clustered, which is generally the case in most channel proteins and neuro-receptors, but the most interesting investigation is the consistent conservation of phenylalanine (F at position 269) and leucine (L at position 353) in all 20 proteins with no mutations. We expect these residues to play some role as a binding centre for interaction of these proteins with several other genes/proteins such as c-jun, mGluR3, Jerky, BDNF, FGF-2, IGF-1, GALR1, NOS and S100beta that are also believed to have a regulatory effect upon these receptors. Therefore, we expect that the gene interaction observations, which we notice from our abstract CNGM, may be justified because of such truly conserved residues. However we must say that all such hypotheses remain unproven and these predictions need to be tested through laboratory experimentation.

For these neuronal proteins we also obtained a dendogram (Fig. 3) that represents the evolutionary related closeness between these subunit proteins. More closely related pairs of neuronal protein sequences have aligned most readily to each other than more divergent pairs.

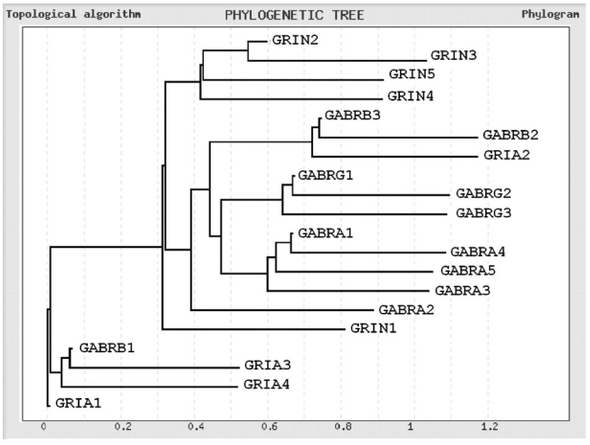
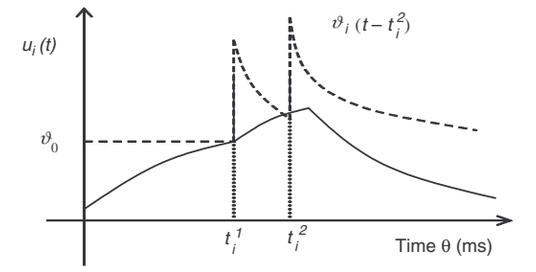


Fig. 3. Phylogenetic tree representing 20 subunit proteins of AMPA, $GABA_A$ and NMDA receptors.

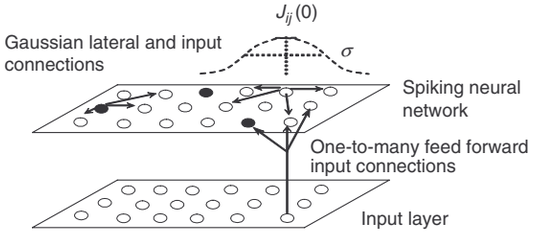
Based on the result of analysis that clearly showed us the extent of conservation of many amino-acid residues among all investigated receptor subunits, we assume that the expression of these individual subunits should be coordinated within one gene group. In addition, these regions can be the basis for mutual interactions. Mutual interaction between subunits of different receptors has been recently confirmed experimentally.[26] Furthermore, this biological insight about our model is also supported by most of similar biological statements available in neuroscience area, especially in case of channel proteins and receptors. Overall interpretation provided us also the motivation for CNGM to treat the subunits of AMPAR, GABRA, NMDAR, and other receptors as a one gene group.

4. CNGM using a Spiking Neural Network

The spiking neural network (SNN) used in our CNGM has been first described in Ref. 4. The SNN model is based upon a classical Spike Response Model (SRM).[27] SRM is a highly simplified spiking model of a neuron and neglects many aspects of neuronal dynamics. In particular, all types of postsynaptic potentials are assumed to have the same shape. However, the shape of postsynaptic potentials depends on the location of the synapse on the dendritic tree. Synapses that are located at the distal end of the dendrite are expected to evoke a smaller postsynaptic response at the soma than a synapse that is located directly on the soma. If several inputs occur on the same dendritic branch within a few milliseconds, the first input will cause local changes of the membrane potential that influence the amplitude of the response to the input spikes that arrive slightly later. Such nonlinear interactions are neglected in the SRM. A purely linear dendrite, on the other hand, can be incorporated into the model.[27] Thus, the state of a neuron i is described by the state variable $u_i(t)$ that can be interpreted as a total somatic postsynaptic potential (PSP) resulting from activation of all synaptic inputs. When $u_i(t)$ reaches the firing threshold $\vartheta_i(t)$ from below, neuron i fires, i.e., emits a spike (see Fig. 4). The moment of $\vartheta_i(t)$ crossing defines the firing time t_i of an output spike. The value



(a)



(b)

Fig. 4. (a) Spiking neuron model. When the membrane potential $u_i(t)$ of the ith spiking neuron reaches the firing threshold $\vartheta_i(t)$ at time t_i^k, the neuron fires an output spike. $\vartheta_i(t)$ rises after each output spike and decays back to the resting value ϑ_0. (b) The SNN architecture. About 10–20% of N = 120 neurons are inhibitory neurons that are randomly positioned on the grid (filled circles).

of the state variable $u_i(t)$ is the weighted sum of all individual PSPs, $\varepsilon_{ij}\big(t-t_j-\Delta_{ij}^{ax}\big)$, such that:

$$u_i(t)=\sum_{j\in\Gamma_i}\sum_{t_j\in F_j}J_{ij}\varepsilon_{ij}\big(t-t_j-\Delta_{ij}^{ax}\big) \tag{4}$$

The weight of synaptic connection from neuron j to neuron i is denoted by J_{ij}. It takes positive (negative) values for excitatory (inhibitory) connections, respectively. Lateral and input connections have weights that decrease in value with distance from neuron i according to a Gaussian formula while the connections themselves can be established at random (with $p=0.5$). Δ_{ij}^{ax} is an axonal delay between neurons i and j which increases with Euclidean distance between neurons. The positive kernel $\varepsilon_{ij}\big(t-t_j-\Delta_{ij}^{ax}\big)=\varepsilon_{ij}(s)$ expresses an individual postsynaptic potential (PSP) evoked on neuron i when a presynaptic neuron j from the pool Γ_i fires at time t_j. For an individual PSP a double exponential formula can be used, i.e.,

$$\varepsilon_{ij}^{type}(s)=A^{type}\left(\exp\left(-\frac{s}{\tau_{decay}^{type}}\right)-\exp\left(-\frac{s}{\tau_{rise}^{type}}\right)\right) \tag{5}$$

where $\tau_{decay/rise}^{type}$ are time constants of the fall and rise of an individual PSP, respectively, A is the PSP's amplitude, and *type* denotes one of the following: *fast_excitation, fast_inhibition, slow_excitation, and slow_inhibition*. These types of PSPs are based on neurobiological data.[28–32] Thus, excitatory as well as inhibitory PSPs have the fast and slow components. Immediately after firing the output spike at t_i, neuron's firing threshold $\vartheta_i(t)$ increases k-times and then returns to its initial value ϑ_0 in an exponential fashion:

$$\vartheta_i(t-t_i)=k\times\vartheta_0\exp\left(-\frac{t-t_i}{\tau_{decay}^{\vartheta}}\right) \tag{6}$$

where τ_{decay}^{ϑ} is the time constant of the threshold decay. In such a way, absolute and relative refractory periods are modelled. External inputs from the input layer are added to the right hand side of Eq. (4) at each time step. Each external input has its own weight $J_i^{ext\_input}$ and an evoked PSP $\varepsilon_i^{fast\_excitation}(t)$, i.e.,

$$u_i^{ext\_input}(t)=J_i^{ext\_input}\,\varepsilon_i^{fast\_excitation}(t) \tag{7}$$

We employed random input with average frequency between 10–20 Hz. Figure 4 illustrates the set up of the SNN.

Table 3 lists value ranges of SNN parameters. These values were taken from neurobiological studies and models and were further adjusted by experimentation so that the resulting SNN activity was neither too low (no spikes) or too high (permanent synchronization) (for the inhibitory and excitatory parameters values see e.g., Refs. 28–32).

5. Using CNGM for Hypothesis Testing and New Discoveries in Brain Study

Our hypothesis is that neural electrical activity with particular spectral characteristics can in fact be achieved through important gene interaction patterns, whereby these patterns can be present in different GRNs. Using CNGM we can test this hypothesis and discover which gene interaction patterns are important for the desired spectral characteristics. We will achieve a desired SNN output through optimization of the model 294 parameters.

Table 3. Model parameters and their value ranges.

Model parameters	Value range
Fast excitation: Amplitude	0.5–3.0
Rise/decay time constants	1–5/5–10
Slow excitation: Amplitude	0.5–4.0
Rise/decay time constants	10–20/30–50
Fast inhibition: Amplitude	4–8
Rise/decay time constants	5–10/20–30
Slow inhibition: Amplitude	5–10
Rise/decay time constants	20–80/50–150
Resting firing threshold,	19–25
Decay time constant/rise after spike	5–50/2–5
Proportion of inhibitory neurons	0.15–0.2
Probability of external input firing	0.011–0.019
Peak/sigma of external input weight	5–10/0.1–2
Peak/sigma of lateral exc weights	5–14/2–8
Peak/sigma of lateral inh weights	10–60/4–10
Unit delay in e/i spike propagation	1 ms/2 ms
Probability of connection	0.5
Number of neurons	120

We are optimizing the interaction matrix W between 16 genes, initial values of neural parameters, architectural parameters of SNN (except the total number of neurons, spike delays and probability of establishing a synaptic connection) and input frequency to the SNN. All model parameters and their value ranges to choose from during optimization are listed in Table 3. We keep the record of spiking activities of all neurons as well as record of the local field potential (LFP). LFP is defined as an average of all instantaneous membrane potentials, i.e., $\Phi(t) = (1/\mathrm{N})\Sigma u_i(t)$. We evaluate the LFP of the SNN by means of FFT (Fast Fourier Transform) in order to compare the SNN output with the EEG signal analysed in the same way.[33] This comparison is based on evidence that brain LFPs in principle have the same spectral characteristics as EEG.[34] Desired signal was real human interictal resting EEG data obtained with permission from Ref. 35. The FFT of the LFP/EEG signal was done for clinically relevant frequency bands. These bands are: delta (0.5–3.5 Hz), theta (3.5–7.5 Hz), alpha (7.5–12.5 Hz), beta 1 (12.5–18 Hz), beta 2 (18–30 Hz), and gamma (>30 Hz).

In order to find an optimal GRN so that the frequency characteristics of the LFP of the associated SNN are similar to the brain EEG characteristics, we use the following simple optimization procedure:

- Generate a population of N CNGMs, each with randomly generated values of coefficients for the GRN matrix **W**, initial gene expression values $\mathbf{g}(0)$, initial values of parameters $\mathbf{P}(0)$ (within the intervals from Table 3), and other model parameters;
- Run each CNGM over a period of time T and record the LFP for each associated SNN;
- Calculate the average spectral characteristics, i.e., the average RIR vectors of the LFP using FFT. RIR for the kth frequency band is defined as: $\mathrm{RIR}_k = I_k/I$, where I_k is the power of the kth frequency band in the signal and I is the total power of the signal. RIRs values change over time therefore we calculate the vector of average RIR values including each frequency band;
- Compare the frequency spectral characteristics of each SNN LFP to the quantitative characteristics of the target EEG signal. Evaluate the closeness of the LFP signal for each SNN to the target EEG signal characteristics using Euclidean distance between their vectors of average RIR values;
- Find CNGM models that match the EEG spectral characteristics better than other solutions.[a] Let us say, the Euclidean distance between the characteristic vectors be smaller than 0.1;
- Analyse the GRN interaction matrices **W** for significant patterns that lead to the desired behaviour.
- Repeat the above steps if necessary.

N = 400 random gene interaction matrices **W**s show almost a uniform distribution of positive and negative interactions between gene groups as can be seen in Fig. 5 but after optimization the interaction matrix is significantly different (Fig. 6).

Among these 400 random solutions, 15 **W**s matrices led to an SNN LFP with spectral characteristics very close to the target EEG signal in terms of Euclidean distance between the characteristic RIR vectors being smaller than 0.1. Visual inspection of these 15 solutions also confirmed that the LFP signal and its spectral characteristics are similar to the target EEG signal. To discover the knowledge, e.g., to find out what

[a]Alternatively, we can use a genetic algorithm to find an optimal solution or solutions, but for illustration of our knowledge discovery method, the above simple optimization procedure will be sufficient.

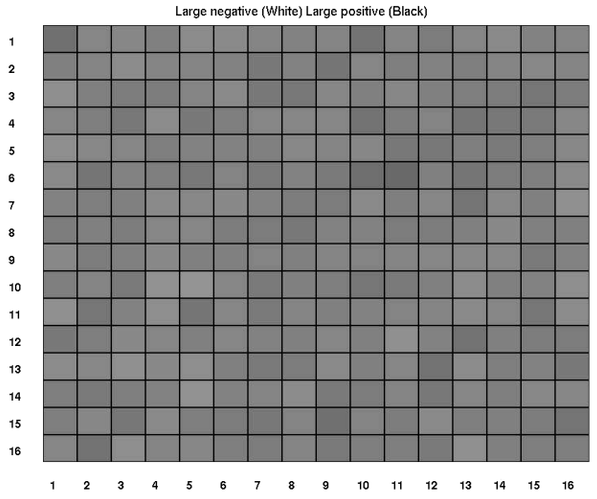


Fig. 5. Intensity of grey reflects upon proportion of positive and negative gene interactions, respectively, in 400 randomly generated interaction matrices **W**, between gene group i (row) and gene group j (column).

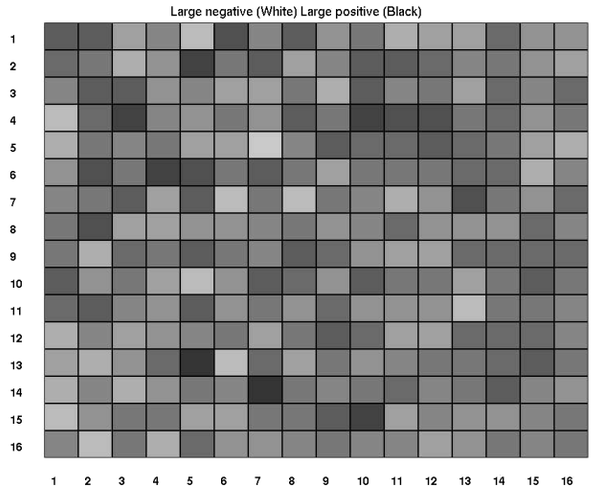


Fig. 6. Intensity of grey reflects upon proportion of positive and negative gene interactions between gene groups i (row) and j (column), respectively, in 15 winner interaction matrices **W** after experimental running of CNGM. In a "winner" GRN the activity of CNGM is similar to the desired brain activity measured as EEG.

these selected solutions have in common, we have calculated how many times the interactions between genes become positive and how many times they become negative (see e.g., Fig. 6). Then we used the χ^2-statistic to make predictions about interactions between gene groups. Let us consider the gene groups, which correspond to excitatory and inhibitory receptors of neurons. We have found that a statistically significant difference at greater than the 0.005 level of probability holds for interactions $w_{4,3} > 0$ and $w_{6,4} > 0$. That is, when the gene group No. 4 (coding for the $GABA_B$ receptor) is elevated in expression, it will be followed by elevated expression of the gene group No. 3 (coding for $GABA_A$ receptor). This interaction means that when there is a need for enhancement of slow inhibition, it will lead also to an increase of the fast inhibition by means of genetic coordination. Second, when the gene group No. 6 (coding for the NMDA receptor) is elevated in expression, it will be followed by elevated expression of gene group No. 4 (coding for $GABA_B$ receptor). This interaction is also very reasonable since one can assume that when the excitation is somehow enhanced, the system tries to keep balance by enhancing inhibition as well, to avoid global synchronizations. Our results suggest that this balancing effect can happen also on the genetic level. Table 4 summarizes all statistically significant deviations from uniformity among all the gene interactions in our sample of 15 **W**s leading to LFP with frequency spectrum similar to an interictal (seizureless) resting EEG.

There are some predictions about gene interactions for genes that are not directly linked to neural parameters. At present these gene groups are not identified and we are not going to concentrate upon them. Most of interactions in Fig. 6 however, were neither positive nor negative on a statistically significant level. In other words, the signs of majority of gene interactions were not crucial for producing desired simulated neural electrical activity. It means that many different GRNs can lead to the same

Table 4. Statistically significant gene interactions from Fig. 6.

χ^2 Significance α	Gene interaction w_{jk}
0.005	$w_{GABRB \rightarrow GABRA} > 0$
	$w_{NMDAR \rightarrow GABRB} > 0$
	$w_{GABRB \rightarrow BDNF} > 0$
	$w_{GABRB \rightarrow FGF\text{-}2} > 0$
	$w_{IGF\text{-}I \rightarrow NMDAR} < 0$
0.025	$w_{AMPAR \rightarrow GALR1} > 0$
	$w_{BDNF \rightarrow IGF\text{-}I} < 0$
	$w_{IGF\text{-}I \rightarrow mGluR3} < 0$
	$w_{IGF\text{-}I \rightarrow GABRA} < 0$
	$w_{S100beta \rightarrow mGlu3} < 0$

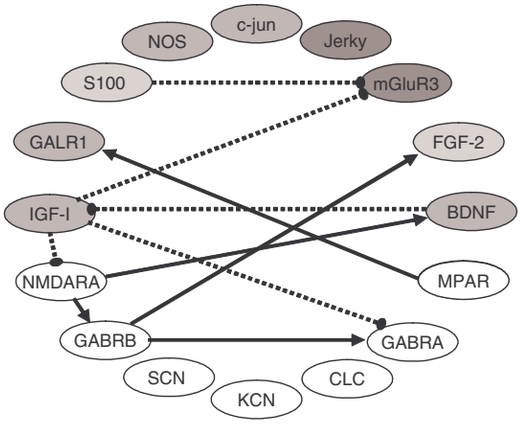


Fig. 7. Illustration of the GRN that resulted from our computer experiments (see Table 4). Solid (dashed) lines denote positive (negative) relationship between genes, respectively. Intensity of grey reflects the level of gene expression in the presented CNGM at a given time moment. Relationship between these genes shown is an abstract one rather than direct physical, chemical or physiological interaction.

observable neural behaviour and that perhaps only few gene interactions are crucial.

Illustration of GRN predicted by our experiments is given in Fig. 7. It should be stressed out that this is an abstract network of gene interactions. Behind each abstract interaction represented by the solid or dashed line in the figure, there is a complex chain of protein-protein interactions leading eventually to up-regulation or down-regulation of the target gene in the network.

As we have learnt now, receptors can have different subunit composition, and still perform the same function, e.g., fast excitation, fast inhibition, etc. However, function parameters like the rise and decay time constants of PSP (opening and closing times of ion channels) will differ for different subunit compositions of receptor-gated ion channels. Indeed, our CNGM has shown that the same behaviour of SNN can be achieved with a range of parameter values (Table 3) corresponding to different receptor subunits compositions, and with different GRNs, corresponding to different interactions between most of the gene groups, provided some important interactions between genes are preserved.

6. Discussion

We have shown, by computational modelling, that the same behaviour of a model neural network and the same characteristics of local field potentials can be achieved with a range of parameter values, as long as some important interactions of gene groups are preserved. Thus our contribution is twofold: (1) the paper shows that behaviour which is the result of interaction between neurons is, on a deeper level, the result of interactions between genes; and (2) it shows that the 'small world assumption' is also valid on the genetic level since only few gene interactions are crucial. Our CNGM was justified by molecular sequence analysis on subunits of crucial neuro-receptors. We have identified highly conserved regions that can serve as the basis for mutual interactions and coordinated expression.

In real neural networks neuronal parameters that define the functioning of a neural network depend on genes and proteins in a complex way. Gene expression values change due to internal dynamics of the gene/protein regulatory network, initial conditions of the genes and external conditions. All this may affect gradually or quickly the functioning of the neural network as a whole. In our computer experiments, we have observed for example that different initial gene values can lead to the same outcome in terms of neuronal activity. Moreover, different types of gene interaction dynamics i.e., be it constant, periodic, quasi-periodic or even chaotic, can lead to a similar LFP of an associated SNN model, provided some statistical distribution of gene interactions is maintained. On the other hand, in the diseased brain, either altered initial conditions expressed in values of neural parameters, mutated genes and/or altered interactions within GRN lead to abnormalities in network activity. Realistic models of gene networks within neural networks should account for these processes.

In order to investigate these phenomena, we have set up a novel model of a CNGM that is simple and biologically plausible. Associated SNN uses principles from the simple spiking neuron models.[27] It should be stressed though that a particular neural network model, and parameters governed by genes, should be chosen based on a problem, which we want to account for. On the example of a resting human EEG signal we have demonstrated our CNGM approach with the introduction of generic CNGM equations, optimization and knowledge discovery techniques. This paper clearly indicates that for obtaining any neurogenetic model with certain accuracy, the integration of

knowledge from different related fields is necessary. Such integration may help in assignment of initial gene/protein expression values for repeating CNGM simulations with improved quality of the **W** matrices results. Knowledge may come from biological lab experiments, bioinformatics databases, molecular sequence analysis, evolving brain-gene ontologies and so on. Domain experts are the one who can analyse and integrate the meanings for simulation of neurogenetic models and make concluding remarks.

Acknowledgements

The work has been supported by the NERF grant AUTX02001 funded by FRST, and by KEDRI funds (www.kedri.info).

References

1. G. Marcus, *The Birth of the Mind: How a Tiny Number of Genes Creates the Complexity of the Human Mind* (Basic Books, New York, 2004).
2. J. L. Holter, A. Humphries, V. Crunelli and D. A. Carter, Optimisation of methods for selecting candidate genes from cDNA array screens: application to rat brain punches and pineal, *J. Neurosci. Methods* **112** (2001) 173–184.
3. The Nervous System, in *Genes and Disease* (National Centre for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call = bv.View..ShowSection&rid = gnd.chapter.75, 2003).
4. N. Kasabov and L. Benuskova, Computational neurogenetics, *Journal of Computational and Theoretical Nanoscience* **1** (2004) 47–61.
5. N. Kasabov and L. Benuskova, Theoretical and computational models for neuro, genetic, and neurogenetic Information processing, in *Handbook of Computational and Theoretical Nanotechnology*, eds. M. Rieth and W. Schommers (American Scientific Publishers, Los Angeles, CA, 2005): Chap. 41.
6. N. Kasabov, L. Benuskova and S. G. Wysoski, Biologically plausible computational neurogenetic models: modeling the interaction between genes, neurons and neural networks, *Journal of Computational and Theoretical Nanoscience* **2** (2005) 569–573.
7. N. Burnashev and A. Rozov, Genomic control of receptor function, *Cellular and Molecular Life Sciences* **57** (2000) 1499–1507.
8. M. Uusi-Oukari, J. Heikkilä, S. T. Sinkkonen, R. Mäkelä, B. Hauer, G. E. Homanics, W. Sieghart, W. Wisden and E. R. Korpi, Long-range interactions in neuronal gene expression: evidence from gene targeting in the $GABA_A$ receptor ß2–α6–α1–γ2 subunit gene cluster, *Molecular and Cellular Neuroscience* **16** (2000) 34–41.
9. S. J. Myers, R. Dingledine and K. Borges, Genetic regulation of glutamate receptor ion channels, *Annu. Rev. Pharmacol.* **39** (1999) 221–241.
10. D. C. Weaver, C. T. Workman and G. D. Stormo, Modeling regulatory networks with weight matrices, in *Pacific Symp. Biocomputing* **4** (World Scientific Publ. Co., Singapore, 1999) pp. 112–123.
11. L. F. A. Wessels, E. P. van Someren and M. J. T. Reinders, A comparison of genetic network models, in *Pacific Symp. Biocomputing* **6** (World Scientific Publ. Co., Singapore, Hawai, 2001) pp. 508–519.
12. N. Kasabov, Z. S. H. Chan, V. Jain, I. Sidorov and D. Dimitrov, Gene regulatory network discovery from time-series gene expression data — a computational intelligence approach, in *Lecture Notes in Computer Science* **3316** eds. N. R. Pal, N. Kasabov, R. K. Mudi, S. Pal & S. K. Parui (Springer-Verlag, Berlin, 2004) pp. 1344–1353.
13. R. Jansen, D. Greenbaum and M. Gerstein, Relating whole-genome expression data with protein-protein interactions, *Genome Research* **12** (2002) 37–46.
14. D. Greenbaum, C. Colangelo, K. Williams and M. Gerstein, Comparing protein abundance and mRNA expression levels on a genomic scale, *Genome Biology* **4** (2003) 117.1–117.8.
15. C. E. M. van Beijsterveldt and G. C. M. van Baal, Twin and family studies of the human electroencephalogram: a review and meta-analysis, *Biological Psychology* **61** (2002) 111–138.
16. G. Buzsaki and A. Draguhn, Neuronal oscillations in cortical networks, *Science* **304** (2004) 1926–1930.
17. B. Porjesz, L. Almasy and H. J. Edenberg, Linkage disequilibrium between the beta frequency of the human EEG and a $GABA_A$ receptor gene locus, in *Proc. Natl. Acad. Sci. USA* **99** (2002), pp. 3729–3733.
18. P. D'Haeseleer, X. Wen, S. Fuhrman and R. Somogyi, Linear modeling of mRNA expression levels during CNS development and injury, in *Pacific Symposium on Biocomputing* **4** (World Scientific Publ. Co., Singapore, 1999), pp. 41–52.
19. M. Simonato, R. Molteni, G. Bregola, A. Muzzolini, M. Piffaneli, L. Beani, G. Racagni and M. Riva, Different patterns of induction of FGF-2, FGF-1 and BDNF mRNAs during kindling epileptogenesis in the rat, *Eur. J. Neurosci.* **10** (1998) 955–963.
20. Y. Li, J. Wang, J. G. Sheng, L. Liu, S. W. Barger, R. A. Jones, L. J. VanEldik, R. E. Mrak and W. S. Griffin, S100 beta increases levels of beta-amyloid precursor protein and its encoding mRNA in rat neuronal cultures, *J. Neurochemistry* **71** (1998) 1421–1428.
21. C. M. Cheng, R. R. Reinhardt, W. H. Lee, G. Joncas, S. C. Patel and C. A. Bondy, Insulin-like growth factor 1 regulates developing brain glucose

metabolism, *Proc. Natl. Acad. Sci. USA* **97** (2000) 10236–10241.
22. M. Toth, J. Grimsby, G. Buszaki and G. P. Donovan, Epileptic seizures caused by inactivation of a novel gene, jerky, related to centromere binding protein-B in transgenic mice, *Nature Genetics* **11** (1995) 71–75.
23. S. Zagulska-Szymczak, R. K. Filipkowski and L. Kaczmarek, Kainate-induced genes in the hippocampus: lessons from expression patterns, *Neurochemistry International* **38** (2001) 485–501.
24. K. Wojtal, A. Gniatkowska-Nowakowska and S. J. Czuczwar, Is nitric oxide involved in the anticonvulsant action of antiepileptic drugs?, *Polish J. Pharmacology* **55** (2003) 535–542.
25. R. Chenna, H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins and J. D. Thompson, Multiple sequence alignment with the Clustal series of programs, *Nucleic Acids Res.* **31** (2003) 3497–3500.
26. V. Salonen, S. Kallinen, F. R. Lopez-Picon, E. R. Korpi, I. E. Holopainen and M. Uusi-Oukari, AMPA/kainate receptor-mediated up-regulation of $GABA_A$ receptor δ subunit mRNA expression in cultured rat cerebellar granule cells is dependent on NMDA receptor activation, *Brain Res.* (2006) in press.
27. W. Gerstner and W. M. Kistler, *Spiking Neuron Models* (Cambridge Univ. Press, Cambridge, MA, 2002).
28. R. A. Deisz, $GABA_B$ receptor-mediated effects in human and rat neocortical neurones *in vitro*, *Neuropharmacology* **38** (1999) 1755–1766.
29. A. Destexhe, Spike-and-wave oscillations based on the properties of $GABA_B$ receptors, *J. Neurosci.* **18** (1998) 9099–9111.
30. F. Wendling, F. Bartolomei, J. J. Bellanger and P. Chauvel, Epileptic fast activity can be explained by a model of impaired GABAergic dendritic inhibition, *Eur. J. Neurosci.* **15** (2002) 1499–1508.
31. J. A. White, M. I. Banks, R. A. Pearce and N. J. Kopell, Networks of interneurons with fast and slow γ-aminobutyric acid type A ($GABA_A$) kinetics provide substrate for mixed gamma-theta rhythm, in *Proc. Natl. Acad. Sci. USA* **97** (2000) 8128–8133.
32. I. C. Kleppe and H. P. C. Robinson, Determining the activation time course of synaptic AMPA receptors from openings of colocalized NMDA receptors, *Biophys. J.* **77** (1999) 1418–1427.
33. R. Q. Quiroga, Quantitative Analysis of EEG Signals: Time-Frequency Methods and Chaos Theory. Thesis, in *Institute of Physiology and Institute of Signal Processing* (Medical University Lübeck, Lübeck, 1998).
34. A. Destexhe, D. Contreras and M. Steriade, Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states, *J. Neuroscience* **19** (1999) 4595–4608.
35. R. Q. Quiroga, *Dataset # 3: Tonic-clonic (Grand Mal) seizures* (http://www.vis.caltech.edu/∼rodri/data.htm, 1998).