Cellular Phenotype Modeling of the Long QT Syndrome Gene Supported by Distributed/Parallel Computation

C Wang[1,2], CD Nugent[1], A Krause[2], W Dubitzky[1]

[1]University of Ulster, Northern Ireland
[2]University of Applied Sciences Wildau, Germany

Abstract

Characterizing the cellular phenotype of the mutated long QT syndrome (LQTS) gene is the bridging approach to relating the molecular dysfunction to the electrophysiological abnormalities. In this study, we model the cellular functional characteristics of the delayed rectifier K+ channel, I_{kr}, responsible for the LQTS subtype 2. The Levenberg-Marquardt algorithm was used to estimate the transition rate function of the Markov model of the channel kinetics, whereby the calculation of the gradient matrices of the merit function was decomposed and implemented by use of the message passing interface. With up to 32 CPUs, the parallel efficiency of the parallel computation was greater than 89%. The kinetic models of the wide-type and mutated wide-type/p.Pro872fs channels were established with a significant improvement in computing efficiency.

1. Introduction

In the exploration of molecular causes of LQTS, a large number of mutations and genetic variants of cardiac ion channels have been identified. The functional properties of specific mutants can be investigated using advanced molecular and microscopic techniques. Thus far, several cellular phenotypes of LQTS mutations have been determined mechanistically, including haploinsufficiency, dominant negative suppression, altered kinetics and trafficking deficiency [1, 2].

With more insights into the varying severity of defective genes, there has been increasing amount of evidence to show a poor correlation of severe mutations with their clinical phenotype manifestations [3]. To clarify such discrepancies, integrative functional studies have been proposed as an alternative. The p.Pro872fs mutation is a heterozygous frameshift mutation recently identified in the C-terminus of the KCNH2, also known as HERG gene. Mechanistically, mutations in this gene have been thought to cause a reduction of the delayed rectifier current and prolongation of action potential duration thus leading to the clinical phenotype LQTS subtype 2 (LQT2). In contrast, p.Pro872fs has been found to have counteracting molecular mechanisms, including both a gain and a loss of function through increased kinetics and reduced intracellular transport [2]. An integrative functional study of such a mutant is appealing in the elucidation of the pathogenesis of LQTS associated lethal cardiac events.

The functional studies reported so far have highlighted that Markov modeling of mutant channel kinetics is a prospect bridging method to link the molecular findings to their electrophysiological properties [4]. Markov modeling is a challenging as well as a computational demanding process, especially when the underlying conformational kinetics of the mutated ion channels are difficult to identify and the analytical derivatives of the merit function required in optimization algorithms are not derivable.

In this paper, we present a new approach to compute the derivates of merit function numerically, independent of an imposition of diagonalization constraint. We chose to model the wide-type (WT) and WT/p.Pro872fs channels. The Levenberg-Marquardt algorithm (LM) was used to estimate the transition rate functions and the calculation of the gradient matrices of the merit function was decomposed and implemented using message passing interface (MPI). In addition to the significantly increased computing efficiency, the representative kinetics of the mutated ion channel in the current study shows a paradoxical functional characterization.

2. Methods

2.1. Data preparation

Kinetic data of activation and inactivation was based on the classic work of Paulussen et al. [2] using transfected HEK293 cells at temperature 20 - 22 °C. The inactivation protocol from Zou's original work [5] was used to verify the inactivating properties of channels. Kinetic data of native I_{kr} of other species was drawn from Tseng's review [6]. Matlab was used to explore graphical data, as which a

large number of kinetic measurements are depicted.

2.2. Distributed/parallel platform

Sun Grid Engine (SGE, version 5.3), was installed and configured on a cluster consisting of nodes with 2785.58 MIPS, 38981.31 MIPS, 3555.23 MIPS and 4734.97 MIPS. The two variants of MPI, i.e., MPICH and LAM/MPI, were implemented and integrated into the SGE system. The actual request on a certain parallel environment was specified in submission script. Similarly, the resources allowed and the limits for CPU running time and memory on each node were also assigned in the submission script. The queue suspend threshold on each node was set to none. The job submission was allowed to be performed on any node within the Grid as well as via the internet from any remote computer. The security was warranted by the generation of SSH private keys and the exchange of public keys among nodes.

2.3. Identification of model topology

The general mathematical representation of any Markov scheme describing ion channel kinetics can be formulated as time-homogeneous system as represented in Equation (1):

$$\frac{dp}{dt} = Q^T p \ , \tag{1}$$

where p is an arbitrary n-element column vector, whose element defines the time evolution of the occupancy probability of each conformation state; Q is the transition rate matrix with its elements represented in the form:

$$Q_{ij} = k_{ij}, i \neq j \ , \tag{2}$$

$$Q_{ii} = \sum_{j \neq i} k_{ij} \ , \tag{3}$$

where k_{ij} is the transition rate from the state i to the state j at the given time t.

The actual dimension of Equation (1) and the estimate of in Equation (2) and Equation (3) were determined by model identification. Briefly, the distribution space of the time constants of exponential components of mathematical models were searched as an alternative to the consideration of transition rates of mathematical models. Transition rate patterns were defined and quasi random seed sequences for each pattern were generated using multiple recursive generator algorithm. A distributed Monte Carlo simulation was performed to investigate various schemes of Markov models and to obtain their inherent distribution space of the time constants. In this study, theoretical Markov schemes were extended to have 10 parameters with 2^{10} transition rate patterns. The underlying conformations and suitable Markov topology was identified by analyzing the probability of experimental kinetic time constants in the distribution space of various schemes.

2.4. Algorithm of gradient decomposition

When Q is diagonalizable, Equation (1) can be derived by means of the matrix exponential as represented in Equation (4):

$$p(t) = e^{Q^T t} p_0 \ , \tag{4}$$

where p_0 is a vector of the initial state occupancy. The current flowing through N ion channels can be formulated as Equation (5):

$$I(V,t) = N(V - V_{rev}) g^T p(t) \ , \tag{5}$$

where g is the conductance vector and V_{rev} the reversal potential. The transition rate between two states is defined as in Equation (6) according to the thermodynamic theory:

$$k_{ij} = e^{(A_{ij} + B_{ij} V_k)} \ . \tag{6}$$

The merit function for the LM nonlinear least squares fit was defined as reprented in Equation (7):

$$f_k = \frac{(I(V_k) - I_k)}{\sigma_k} \ , \tag{7}$$

where σ_k is independent Gaussian error. Nevertheless, the analytical form of Equation (4) can not be derived when Q is not diagonalizable. More importantly, the derivative information with respect to specific parameters (A_{ij}, B_{ij}) would be often missing in the eigenvectors of matrix Q, even in the case of a diagonalizable matrix Q. Let's set e^{tQ^T} be:

$$e^{tQ^T} = I + \sum_{i=1}^{\infty} \frac{t^i}{i!} {Q^T}^i \ , \tag{8}$$

where I is an $n \times n$ unit matrix. The derivatives of the structure matrix Q with respect to k_{ij}can be derived as:

$$\frac{\partial Q}{\partial k_{ij}} = S^{ij} - S^{ii} \ , \tag{9}$$

where S^{ij} and S^{ii} are single entry matrices which are zero everywhere except in the entry (i,j) or (i,i). Subsequently, the derivatives of a matrix product can be obtained in the form:

$$\frac{\partial {Q^T}^2}{\partial k_{ij}} = Q^T (S^{ij} - S^{ii}) + (S^{ij} - S^{ii}) Q^T \ . \tag{10}$$

Assume that the discrete time step spectrum used in the numerical integration is $t_0, ..., t_{k-1}, t_k, ..., t_N$ and the following condition is fulfilled:

$$t = \sum_{k=0}^{N} t_k \ , \tag{11}$$

then, the derivative of the kth element of $p(t)$ in Equation (4) with respect to the parameter k_{ij} can be approximated by recursively computing Equation (12):

$$\frac{\partial p^T{}_{t_{k+1}}}{\partial k_{ij}} = \frac{\partial p^T{}_{t_k}}{\partial k_{ij}} e^{t_k Q^T} + p^T{}_{t_k} e^{t_k Q^T} (S^{ij} - S^{ii}) \ . \quad (12)$$

Consequently, the approximation of the gradient of Equation (7) can be also obtained numerically. Equations (9) - (12) were implemented using MPI function library. This parallelism is dependent on the amount of the rates to be solved in the nonlinear fitting.

3. Results

3.1. Performance of parallel computation

CardioWave [7] was deployed and run on the cluster to measure the computing efficiency. For a 2-D single membrane model simulation (316x316 nodes), a speedup of

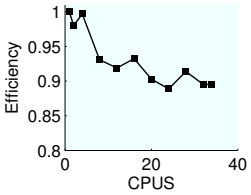

Figure 1. Efficiency of distributed/parallel system

18.62 on 20 processors was achieved. With up to 32 CPUs, the parallel efficiency of our system was greater than 89%, as shown in Figure 1.

3.2. Efficiency and accuracy of algorithm

A five state Markov scheme was used to verify the algorithm of decomposition. The average running time for an iteration was 242.10±2.452 ms on the node with 3555.23 MIPS and 169.74±3.893 ms on the node with 4734.97 MIPS. The computing performance of the gradient decomposition was summarized in the Table 1. The value of

Table 1. Performance of decomposition.

No. of	Max. load		Min. load	
nodes	CPU (μs)	P	CPU (μs)	P
1	105.8±3.09	9	105.8±3.09	9
2	72.5±4.91	5	48.3±0.58	4
3	49.2±4.71	3	28.6±0.58	3
4	50.4±6.44	3	27.0±1.67	2
5-8	37.0±3.40	2	12.9±0.41	1
9	24.8±4.80	1	13.6±2.87	1

P indicates the amount of transition rates for which the derivates must be calculated on a node. As seen from the forth column, a speedup of 7.78 was achieved with 9 nodes. Comparatively, the second column shows a lower speedup. The reason for the phenomenon is that the non-parallel parts of programs were performed on the first node which had a maximal computation load.

A simplified form of Equation (1) with two variables and three parameters was used to verify the accuracy of the algorithm. The results are shown in Figure 2. The analytical

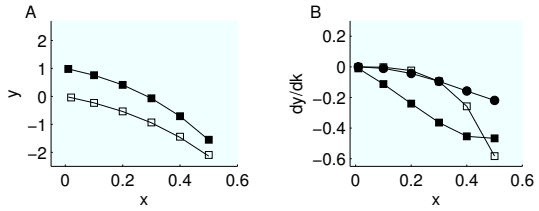

Figure 2. Accuracy verification of algorithm. (A) Numerical values of y_1 (filled square) and y_2 (open square); analytical solutions of y_1 and y_2 (solid line). (B) Derivatives with respect to k_1 (filled square), k_2 (open square) and k_3 (filled circle).

values of parameters are 2, 6 and 5 respectively, and the estimates obtained by our algorithm were 2.004±0.0075, 5.931±0.1034 and 5.110±0.0957 respectively.

3.3. Model decision

Model choice and initial parameter estimation were performed as described in the section 2.2. The time constant distributions of two Markov schemes are shown in Figure 3. Compared to the scheme shown in Figure 3 (B), the scheme in Figure 3 (A) has a wider distribution. The data suggests that the appropriate scheme could represent acti-

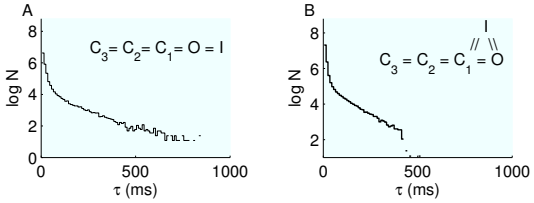

Figure 3. Time constant distributions of Markov schemes.

vating kinetics of the WT/p.Pro872fs channel, whose time constants for -20 mV to 10 mV are much larger than 500 ms.

3.4. Characterization of channel kinetics

The Markov models of WT and co-expression WT/p.Pro872fs channels were constructed and their transition rates have been listed in Table 2, where k_{ij} refers to

Table 2. WT and WT/p.Pro872fs transition rate functions.

	KCNH2	KCNH2 + p.Pro872fs
k_{32}	$e^{(-1.598+2.460E-3V)}$	$e^{(-1.423+4.630E-3V)}$
k_{23}	$e^{(1.702-8.141E-3V)}$	$e^{(1.3092-2.940E-3V)}$
k_{21}	$e^{(-1.419+7.411E-3V)}$	$e^{(-1.166+5.20E-4V)}$
k_{12}	$e^{(1.434-1.851E-2V)}$	$e^{(1.007-1.428E-2V)}$
k_{1o}	$e^{(-1.120+1.378E-2V)}$	$e^{(-0.971+1.214E-2V)}$
k_{o1}	$e^{(-10.593+7.452E-2V)}$	$e^{(-11.715+5.025E-2V)}$
k_{oi}	$e^{(0.218-2.314E-2V)}$	$e^{(0.876-1.293E-2V)}$
k_{io}	$e^{(-0.761+9.143E-3V)}$	$e^{(-1.129+1.984E-2V)}$

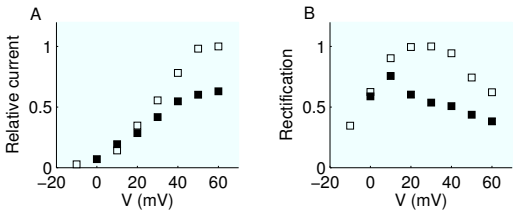

Figure 4. Activation and rectification. WT (filled square) WT/p.Pro872fs (open square).

the transition rate between two conformation states. The activation and rectification characteristics predicted by the models are shown in Figure 4. In addition with a paradoxical phenomenon in the activation, the WT/p.Pro872fs showed a positive shift in the rectification.

4. Discussion and conclusions

The advantage of SGE integrated with MPI parallel environment over a pure parallel computing is its flexibility in resource management and its adaptability to various applications. In this study, both CardioWave [7] simulation and gradient decomposition demonstrated high performance of parallelism in computation-intensive application. In contrast, Monte Carlo simulation for searching the probability distribution functions of Markov schemes was partitioned in the form of array-jobs in SGE, in combination with a partially parallelization for the evaluation function. This approach demonstrated the flexibility of such a platform in solving either parallel-dominating or distribution-dominating computation.

The mutation p.Pro872fs was experimentally investigated by Paulussen et al. [2] recently using heterologous expression (HEK293) and microscopic techniques. Their results show that the mutant may lead to cellular dysfunction through both increased kinetics and reduced intracellular transport [2]. Our model predication demonstrated that the co-expression WT/p.Pro872fs decreased the amplitude of expressed current in the low voltage range, but increased the amplitude significantly in the high voltage range. While a loss of function of is the cause of LQT2, a gain of function in the channel would cause a shorten QT interval rather than a prolonged QT interval. Paulussen et al. [2] suggested a decreasing in trafficking deficiency of heteromultimeric channel has superseded the gain of function by kinetic increase, thus leading to LQT2. As shown in Figure 4, for the membrane potentials (-20 mV to +20 mV) typical of the plateau phase of cardiac action potentials, a kinetic decrease in repolarizing power may still play a role in lengthening QT interval. Whether the interplay of these two molecular mechanisms lead to the severe long QT phenotype need to be proved in a further study.

In conclusion, our data has shown that the distributed/parallel computation is an effective and flexible low-cost measure to solve problems requiring high computational power. Benefiting from this platform the kinetic models of both WT and WT/p.Pro872fs channels have been established. To our knowledge, this is the first attempt to use Markov model to characterize the kinetics of a C-terminus frameshift mutation found in the LQT2 gene carriers.

References

[1] Moss AJ, Kass RS. Long QT syndrome: from channels to cardiac arrhythmias. J Clin Invest 2005;115(8):2018–2024.

[2] Paulussen A, Raes A, Jongbloed RJ, Gilissen RA, Wilde AA, Snyders DJ, et al. HERG mutation predicts short QT based on channel kinetics but causes long QT by heterotetrameric trafficking deficiency. Cardiovas Res 2005;67(3):467–475.

[3] Bianchi L, Priori SG, Napolitano C, Surewicz KA, Dennis AT, Memmi M, et al. Mechanisms of iks suppression in LQT1 mutants. Am J Physiol Heart Circ Physiol 2000; 279:H3003–H3011.

[4] Clancy CE, Rudy Y. Na+ channel mutation that causes both Brugada and long-QT syndrome phenotypes. Circulation 2002;105(10):1208–1213.

[5] Zou A, Xu QP, Sanguinetti MC. A mutation in the pore region of HERG K+ channels expressed in xenopus oocytes reduces rectification by shifting the voltage dependence of inactivation. J Physiol 1998;509(1):129–137.

[6] Tseng G. Ikr: The hERG channel. J Mol Cell Cardiol 2001; 33(5):835–849.

[7] Pormann JB, Henriquez CS, Board JA, Rose DJ, Harrild DM, Henriquez AP. Computer simulations of cardiac electrophysiology. In Proceedings of the 2000 ACM/IEEE Conference on Supercomputing. Washington, DC: IEEE Computer Society, 2000; 24.

Address for correspondence:

Chong Wang
Universtiy of Applied Sciences Wildau,
Bahnhofstr. 1, 15745 Wildau, Germany
cwang@igw.tfh-wildau.de