Idealizing Ion Channel Recordings by a Jump Segmentation Multiresolution Filter

Thomas Hotz*, Ole M. Schütte, Hannes Sieling, Tatjana Polupanow, Ulf Diederichsen, Claudia Steinem, and Axel Munk

***Abstract*—Based on a combination of jump segmentation and statistical multiresolution analysis for dependent data, a new approach called J-SMURF to idealize ion channel recordings has been developed. It is model-free in the sense that no a-priori assumptions about the channel's characteristics have to be made; it thus complements existing methods which assume a model for the channel's dynamics, like hidden Markov models. The method accounts for the effect of an analog filter being applied before the data analysis, which results in colored noise, by adapting existing muliresolution statistics to this situation. J-SMURF's ability to denoise the signal without missing events even when the signal-to-noise ratio is low is demonstrated on simulations as well as on ion current traces obtained from gramicidin A channels reconstituted into solvent-free planar membranes. When analyzing a newly synthesized acylated system of a fatty acid modified gramicidin channel, we are able to give statistical evidence for unknown gating characteristics such as subgating.**

***Index Terms*—Event detection, gramicidin derivative, planar patch clamp, reconstruction, statistical multiresolution criterion, subconductance.**

I. Introduction

THE investigation of ion channel functionalities is of particular importance as they play major roles in cellular processes like signal transduction, energy conversion, and transporting [1]. Being involved in many human diseases such as epilepsy, cardiac arrythmias, and others, they are major targets of pharmaceutical drugs [2], [3]. The channel's behavior with respect to ion conductivity and dynamic properties is best monitored on a single channel level, but tedious data analysis is required to gain statistically profound information about channel characteristics and the effect of external stimuli on the channel [4]. Automatic, software based analysis, especially of unknown ion channels, is hampered by a wide dynamic range of channel kinetics with channel opening times from milliseconds up to tens of seconds, rendering this a statistical multiresolution problem. Furthermore the electric conductivity of different ion channels can span several orders of magnitude from picosiemens to tens of nanosiemens [4].

Here, we focus on the problem of idealization, as it is called in this field, of ion channel current traces, i.e., on the reconstruction or estimation of the channel's conductivity over time without noise; one may also call that denoising or signal detection. This is an important step in the analysis of an ion channel's traces as many of its characteristics can be decuced from idealized traces: number of states, open and closed times, transition rates between states [5], or the Nernst potential [6].

Many idealization methods are based on specific models for the channel's behavior [7], e.g., that there is a Markov chain of states, each with its associated conductance. While model-based approaches often have the advantage that they immediately give rise to estimates of the channel's dynamical parameters, they have the drawback that they require a model for the dynamics to be specified in advance, i.e., they assume prior knowledge about the channel's behavior, the results often depending crucially on the model being assumed. This is in particular problematic if one wants to investigate whether the model is appropriate for that channel, e.g., when one doubts the channel behaving according to a Markovian model [8]–[10].

In contrast, methods which only make minimal assumptions about the signal and its dynamics are often labelled "model free"; in statistics, these are called nonparametric methods. They are clearly more appropriate when little is known in advance about the channel as they allow for an unprejudiced analysis of the data. The aim of this research was to develop such an idealization technique which is applicable in the general situation where nothing is known about a channel.

Currently, to the best of our knowledge there are two popular types of model-free approaches for idealizing ion channel recordings: amplitude thresholding [11], [12] defines a channel to be open if its conductance exceeds a certain, predetermined threshold, otherwise it is assumed to be closed; generalizations to several states or multiple channels are straight-forward. The applicability of this method is adversely affected by a low signal-to-noise ratio (SNR), thus often requiring the data to be filtered with some digital low-pass filter before applying the amplitude threshold.

The slope thresholding approach focuses on the detection of jumps: first, local estimates of the slope, e.g., differences between consecutive observations, are computed; if one exceeds a certain threshold in absolute value, a jump has been detected

Manuscript received May 08, 2013; accepted September 19, 2013. Date of publication November 05, 2013; date of current version January 07, 2014. *Asterisk indicates corresponding author.*

*T. Hotz is with the Institute of Mathematics, Technische Universität Ilmenau, 98693 Ilmenau, Germany (e-mail thomas.hotz@tu-ilmenau.de).

O. M. Schütte, T. Polupanow, U. Diederichsen and C. Steinem are with the Institute of Organic and Biomolecular Chemistry, Georgia Augusta University of Goettingen, 37073 Goettingen, Germany.

H. Sieling is with the Institute for Mathematical Stochastics, Georgia Augusta University of Goettingen, 37073 Goettingen, Germany.

A. Munk is with the Institute for Mathematical Stochastics, Georgia Augusta University of Goettingen, 37073 Goettingen, Germany, and also with the Max Planck Institute for Biophysical Chemistry, 37077 Goettingen, Germany.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNB.2013.2284063

[13], [14]. Again, one often applies a low-pass filter beforehand to reduce the noise. Here, the threshold depends on the noise level and does not have to be chosen by the analyst; in particular, no assumption on the number of conductance levels is required. Another advantage over amplitude thresholding is that slope thresholding is more robust against baseline drifts.

However, both amplitude and slope thresholding usually require low-pass filtering. While additional filtering with a prespecified filter removes noise, thus alleviating the problem of detecting spurious jumps caused by a low SNR, it also affects the true signal by smoothing it out, which in turn may lead to events being missed and the localization precision of jumps deteriorating [11], [12], [15]. Albeit there may be a theoretical optimum for the filter length for detecting events of a given length, i.e., at a fixed scale, this length is not known a priori, nor do events share the same length. In fact, if the channel's open and closed states form a simple Markov chain as in the simplest, realistic model then the event lengths will be exponentially distributed, i.e., event lengths ranging over about three orders of magnitude will occur, and short events will be smoothed out while long events get smoothed less than possible if the filter is adapted to the average event length. Therefore, we propose to use a multiresolution approach which automatically will adapt to the right scale, i.e., event length, in order to smooth as much as possible depending on the event's length, without loosing details for short events. This combines a statistical multiresolution analysis (SMRA) with the specific structure of ion channel recordings, namely that they essentially are piecewise-constant block signals, observed after an analog low-filter has been applied which results in additive colored noise. The signal is determined by the time points at which jumps occur, and by the conductance in between. The SMRA introduced below has been adapted to this situation and will allow to segment the signal at the jump points as it detects with high probability if a jump has been missed. The method which is thereby obtained will thus be called "Jump-Segmentation by MUltiResolution Filter (J-SMURF)".

Before we describe J-SMURF in detail, we state which criteria we would like a model-free idealization method to fulfill, cf. [14]: first of all, the method should not require any a-priori information, in particular no model for the channel's dynamics, or any assumptions about its conductance or the event lengths, thus being universally applicable and hence maximizing objectivity by requiring no subjective choices to be made by the analyst; secondly, while the idealization should be perfect in the absence of noise, the amount of spurious jumps should be controllable statistically.

The latter requirement asks us to take multiple testing into account, i.e., the problem caused by simultaneously looking for jumps in the signal at many locations. If the error of detecting any false-positive, i.e., any spurious jump, is required to be small, this will force threshold values to increase, thereby generally also increasing the number of missed events. Clearly, this issue aggravates the problems described above, calling for an approach that takes all information available from the data into account.

To validate J-SMURF we used simulated data as well as current traces of gramicidin A (gA) and a novel gA derivative. Gramicidin A is a small peptide capable of forming ion channels by a diffusion regulated interleaflet dimerization [16], [17]. Its channel characteristics have been extensively studied with respect to its ion selectivity, dynamic behavior, and the mechanical properties of the surrounding membrane [18]–[23]. Due to its simple structure and small size, gA derivatives with altered channel properties are readily available from chemical synthesis, making it a suitable system to benchmark the idealization approach presented in this work [24].

The results of J-SMURF were compared to results obtained using a slope thresholding method. In particular, we chose the TRANSIT algorithm [14] for the comparison as it has also been specifically developed for the purpose of model-free idealization of ion channel recordings, in particular in case of multiple conductance levels, possibly caused by subconductance states. However, we expect other slope thresholding approaches to exhibit a qualitatively similar behavior. For comparisons with other methods, our implementation of the algorithms described below is available from http://stochastik.math.uni-goettingen.de/smuce.

II. Theory

A. Statistical Multiresolution Analysis

In order to more clearly expose the main ideas, we will for the moment ignore that a low-pass filter has been applied prior to digitization. Considering the simpler model where the data $y_n, n = 1, \ldots, N$, observed at times t_n are given as the true signal $f(t_n)$ plus Gaussian white noise ϵ_n, i.e., we assume:

$$y_n = f(t_n) + \epsilon_n, \tag{1}$$

where the ϵ_n are independently distributed Gaussian random variables with mean 0 and variance σ^2; for the time being, we assume σ^2 to be known. The crucial assumption that we make is that the true signal f is a piecewise constant function; i.e., we assume that the unknown mean of the measured conductance always is constant for some time, then it may jump to a different conductance level at which it remains constant again for a while, and so on. In particular, we assume that the change between conductance levels happens faster than the rate at which we sample, a realistic assumption for many channels [11].

In order to reconstruct the signal f, in a first step we require a statistical multiresolution constraint (SMC) to be fulfilled, cf. e.g. [25]–[29], which will be adapted to the problem of a low pass filter later on in Section II-C. So, assume we were given some candidate, piecewise constant reconstruction g and we had to decide whether this g is a valid reconstruction of the true (unknown) f in (1). The idea of the SMC is to check on a statistically sound basis whether the residuals $y_n - g(t_n)$ indeed behave like white noise *on all levels of resolution, i.e., scales, simultaneously*, as we would expect them to if indeed g was equal to f. For this, consider for each interval $[i, j] = \{i, i+1, \ldots, j\}, 1 \le i \le j \le N$, the statistic:

$$S_{i,j} = \frac{1}{\sigma\sqrt{j-i+1}} \sum_{n=i}^{j} (y_n - g(t_n)). \tag{2}$$

Each of these statistics has mean 0 and variance 1 under the hypothesis that $g = f$. However, if $g \neq f$ there will be some

interval $[i, j]$ on which they differ, say $f(t_n) > g(t_n)$ for $n \in [i, j]$, such that $S_{i,j}$ will have mean:

$$\mathbf{E}S_{i,j} = \frac{1}{\sigma\sqrt{j-i+1}} \sum_{n=i}^{j} (f(t_n) - g(t_n)) > 0; \qquad (3)$$

similarly, $\mathbf{E}S_{i,j} < 0$ in case $f(t_n) < g(t_n)$, such that $|\mathbf{E}S_{i,j}| > 0$ on these intervals while $|\mathbf{E}S_{i,j}| = 0$ if the reconstruction is perfect. As we want to simultaneously test for deviations from the hypothesis on all intervals $[i, j]$, we consider the maximum over all $|S_{i,j}|$, i.e., we construct the multiresolution statistic M, cf. again [25]–[29],

$$M = M(g) = \max_{1 \le i \le j \le N} |S_{i,j}|, \qquad (4)$$

expecting M to be small if $g = f$ and large otherwise.

Indeed, under the hypothesis that $g = f$, the distribution of $M = M(g) = M(f)$ does not depend on f, so we can a priori determine its quantile q such that the probability of $M > q$ is less than some prespecified significance level α, say $\alpha = 5\%$. Put the other way around, a solution g is feasible if $M(g) \le q$, i.e., if it satisfies the SMC at level α. In practice, q will be determined by simulation. We then reject a proposed reconstruction g if its multiresolution statistic $M(g)$ is larger than q. Thereby, we will be able to detect if g has less jumps than the true f, at least if the cumulated difference between f and g in (3) is large enough on some interval.

B. Jump Segmentation

Of course, one cannot detect whether g contains too many jumps with this approach. In fact, if g is allowed to have $N-1$ jumps it can interpolate the data perfectly, $g(t_n) = y_n$, resulting in $M(g) = 0$. Therefore, following ideas of [29] and [30], we determine the simplest reconstruction g in the class of piecewise constant functions, i.e., one with the minimal number of jumps, fulfilling the multiresolution criterion $M(g) \le q$. Here, g may be any piecewise constant function with a finite though arbitrary number of jumps, with no restriction on the blocks' levels. Thus, denoting by G_k the set of piecewise constant functions with at most k jumps, $k \ge 0$, we first determine the minimal number K of jumps necessary to fulfill the criterion, i.e., K is the minimal k for which there is some $g \in G_k$ such that $M(g) \le q$. This g will in general not be unique, so we choose the one which fits the data best in the least-squares sense, i.e., the one which maximizes the likelihood of the data. The estimator $\hat{f}$ is thus given by:

$$\hat{f} = \operatorname*{argmin}_{g \in G_K, M(g) \le q} \sum_{n=1}^{N} (y_n - g(t_n))^2. \qquad (5)$$

By construction of $\hat{f}$, the probability of overestimating the number of jumps is bounded by α. It can be shown that, when σ tends to zero, the multiresolution constraints tighten such that $\hat{f}(t_n)$ tends to $f(t_n)$, while K converges to the true number of jumps; hence, the estimator will reconstruct perfectly if no noise is present. In [29] extensive theory for a scale-calibrated version of (5) has been developed. There, scale-calibration has been used, in order to prevent domination of small scales [26]. As we will argue in the following subsection small scale domination will be automatically excluded due to the low pass filtering of the ion channel recordings and hence scale-calibration on small scales becomes dispensible. On the one hand this simplifies the theory, on the other hand the low pass filtering will complicate the computation of $\hat{f}$ in (5), see Section II-C. In summary, see [29]:

> Under mild assumptions the estimated number of jumps converges to the true number of jumps in probability if the sampling interval goes to zero while the measurement time remains fixed [29, Theorem 2.3]. Furthermore, vanishing signals can be reconstructed at the optimal detection rate [29, Theorem 2.7], and all jump locations are recovered simultaneously with an error which is (asymptotically) proportional to the sampling interval's length [29, Theorem 2.8].

More precisely, non-asymptotic exponential deviation bounds were proved from which these asymptotic statements follow.

Minimizing the number of jumps in the regression context has also been advocated by [31] and [32], with the notable difference, however, that a *global* penalty enforces the solution to be parsimonious which is quite different from the local MRC in (4).

C. Accounting for the Effect of an Analog Low-Pass Filter

We will now discuss how the general approach described above can be adapted to the situation where an analog low-pass filter has been applied prior to digitization, at the same time ensuring that it can efficiently be computed. For the former, we assume that the analog filter's kernel F, i.e., its impulse response function, has compact support, i.e., it is zero outside a finite interval of length L. Moreover, assume that the signal has been uniformly sampled at rate $1/\Delta$, such that $t_n = n\Delta$. Then, if the noise was white and Gaussian before filtering, which is a realistic assumption for ion channel recordings, the observations fulfill:

$$Y_n = (F * f)(t_n) + \tilde{\epsilon}_n, \qquad (6)$$

with $(F * f)(t) = \int F(s) f(t-s)\, ds$ denoting the convolution of f with F, cf. (1). Here the noise $\tilde{\epsilon}_n$ is still Gaussian with mean zero, but it is correlated, i.e., colored, noise now; its standard deviation after filtering will be denoted by $\tilde{\sigma}$. We observe that we at least know $\tilde{\epsilon}_n$ to be independent of $\tilde{\epsilon}_m$ if $|t_n - t_m| > L$. Similarly, if f is piecewise constant for some time $T \ge L$, $F * f$ will be piecewise constant at least on an interval of length $T - L$; shorter events will be difficult to reconstruct. We thus restrict ourselves to only considering those intervals in (4) at which g is piecewise constant for at least length L, then checking whether the observed signal appears to be constant on the interval shortened by L, as $f * g$ would be. To simplify notation, we assume that $F(t) = 0$ for all $t \notin (0, L)$ while $\int F(t)dt = 1$. We then substitute (4) by

$$\tilde{M}(g) = \max_{[i,j] \in I(g)} \left| \frac{1}{\tilde{\sigma}\sqrt{j-i+1}} \sum_{n=i}^{j} (y_n - (F * g)(t_n)) \right|, \qquad (7)$$

where $I(g)$ denotes the set of intervals $[i, j]$ for which g is constant on $[t_i, t_j + L]$, and hence $F * g$ is constant on $[t_i, t_j]$. Proceeding as in (5) before, we finally obtain the jump segmentation multiresolution filter (J-SMURF),

$$\tilde{f} = \underset{g \in G_K, \tilde{M}(g) \leq q}{\operatorname{argmin}} \sum_{n=1}^{N} (y_n - F * g(t_n))^2. \tag{8}$$

With this adaptation, the distribution of $\tilde{M}(g)$ under the hypothesis $g = f$ still depends on f. However, it can be bounded by the distribution of $\tilde{M}(0)$ which does not depend on f anymore:

Theorem 1: Assume model (6) with filter function F and let $\tilde{M}$ denote the multiresolution statistic in (7) underlying J-SMURF (8). Then the distribution of $\tilde{M}(g)$ under the hypothesis $g = f$ can be bounded by that of $\tilde{M}(0)$ under the hypothesis $f = 0$, i.e.,

$$\mathbf{P}_{f=g}(\tilde{M}(g) \geq q) \leq \mathbf{P}_{f=0}(\tilde{M}(0) \geq q) \tag{9}$$

for all $q > 0$.

Proof: For any g, one considers the maximum over coefficients of the form

$$\tilde{C}_{ij} = \left| (\tilde{\sigma}\sqrt{j - i + 1})^{-1} \sum_{n=i}^{j} \tilde{\epsilon}_n \right|. \tag{10}$$

If g contains jumps, $I(g)$ contains fewer intervals than $I(0)$, so that the maximum will be smaller, i.e., $\tilde{M}(g) \leq \tilde{M}(0)$ almost surely, which implies the assertion. ■

This means in particular that the probability of detecting more jumps than present in f can still be controlled to be at most α, if q is chosen such that $P(\tilde{M}(0) \geq q) \leq \alpha$. Therefore, all jumps detected by J-SMURF are statistically significant at error level α.

Note that the reconstruction will still be perfect if no noise is present. In contrast to the situation with white noise above, the coefficients $\tilde{C}_{ij}$ in (10) behave differently for short intervals. Indeed, if the filter F is smooth, as in practice, then the noise $\tilde{\epsilon}$ converges to a smooth random process when the sampling rate goes to infinity while the measurement time remains fixed. Therefore, penalization of short intervals as in [26], [29] is unnecessary, as the maximum over the coefficients, $\tilde{M}$, will already stay (stochastically) bounded without any correction terms. However, this also means that the amount of information within a fixed measurement time is also limited, so one cannot expect the noise to be removed perfectly. Precise mathematical statements about J-SMURF's ability to resolve signals are technically involved and beyond the scope of this paper. Nonetheless, by simulating $\tilde{M}(0)$, we can control the error of detecting spurious jumps so that the reconstructed jumps will be statistically significant, cf. Theorem 1. In the simulations and experimental data examples below, we will demonstrate J-SMURF's precision in practice.

D. Computational Issues

The choice of $I(g)$ allows to efficiently compute the J-SMURF estimator $\hat{f}$ by dynamic programming which can be shown to have a maximum complexity of order $O((N - L)^2)$, see also [30], [33], and [29]. Moreover, following the arguments in [29, Sec. 3] it can be shown that the computation time depends on the final solution and decreases significantly as the final solution includes more jumps. In particular, for ion channel recordings with a large channel activity the computation of J-SMURF will be significantly faster.

E. Estimating the Variance and Final Idealization

Finally, we need to address the problem of estimating $\tilde{\sigma}$. A robust estimator for it is given by dividing the median of the absolute differences $|y_{n+\delta} - y_n|$, δ denoting the smallest integer larger than L/Δ, by $\sqrt{2}$ times the upper quartile of the standard normal distribution [25]. In fact, $y_{n+\delta} - y_n$ is a mean-zero Gaussian random variable with variance $2\tilde{\sigma}^2$, hence the proposed estimator is consistent for a fixed, finite number of jumps in f if $N \to \infty$; for finite N it slightly overestimates $\tilde{\sigma}$, leading to a somewhat conservative multiresolution criterion. Other estimates for $\tilde{\sigma}$ in this situation can be found in [34] and the references given there.

One might view J-SMURF as being like a filter which filters more, if events are longer but without losing detail, in particular not smoothing over the jumps. As a matter of fact, we found it to be so sensitive that it even picks up small movements of the baseline which cannot be avoided in the experiment. However, this is only a minor issue which can be dealt with easily; a simple remedy for it is e.g., to define amplitude thresholds based on the J-SMURF estimator in order to obtain the final idealization. This approach worked well for the examples discussed below but may be replaced by a more sophisticated, fully automatic method if required.

III. Methods

A. Simulations

Adapting the methodology proposed in [14] to the experimental data's parameters, simulations were carried out by first generating a continuous time Markov chain with the corresponding rates for 60 s, or a constant signal in the noise-only case; the signal was tenfold oversampled at 100 kHz, Gaussian white noise was added, and a digital four-pole Bessel low-pass filter at 1 kHz cutoff-frequency applied; finally the signal was subsampled at 10 kHz, resulting in 600,000 data points for each simulation.

B. Experimentals

Gramicidin A was purchased from Sigma-Aldrich (Schnelldorf, Germany) and used without further purification.

The gA derivative was synthesized by solid phase peptide synthesis. The amino acid sequence was derived from gA substituting D-Leu_{12} and D-Leu_{14} by D-Ser(decanoyl), respectively [35]. The oligomer was purified by reverse phase high-performance liquid chromatography and characterized by high resolution mass spectrometry. Giant-unilamellar vesicles (GUVs) (1,2-diphytanoyl-*sn*-glycero-3-phosphocholine/cholesterol, 9:1) containing peptide (nominal lipid to peptide ratio of 10000:1 up to 1000:1) were produced by the electroformation method [36]. Electrical recordings of solvent-free lipid bilayers in buffer (1 M KCl, 10 mM HEPES, pH 7.4, 20°C) were performed using the Port-a-Patch (Nanion Technologies GmbH,

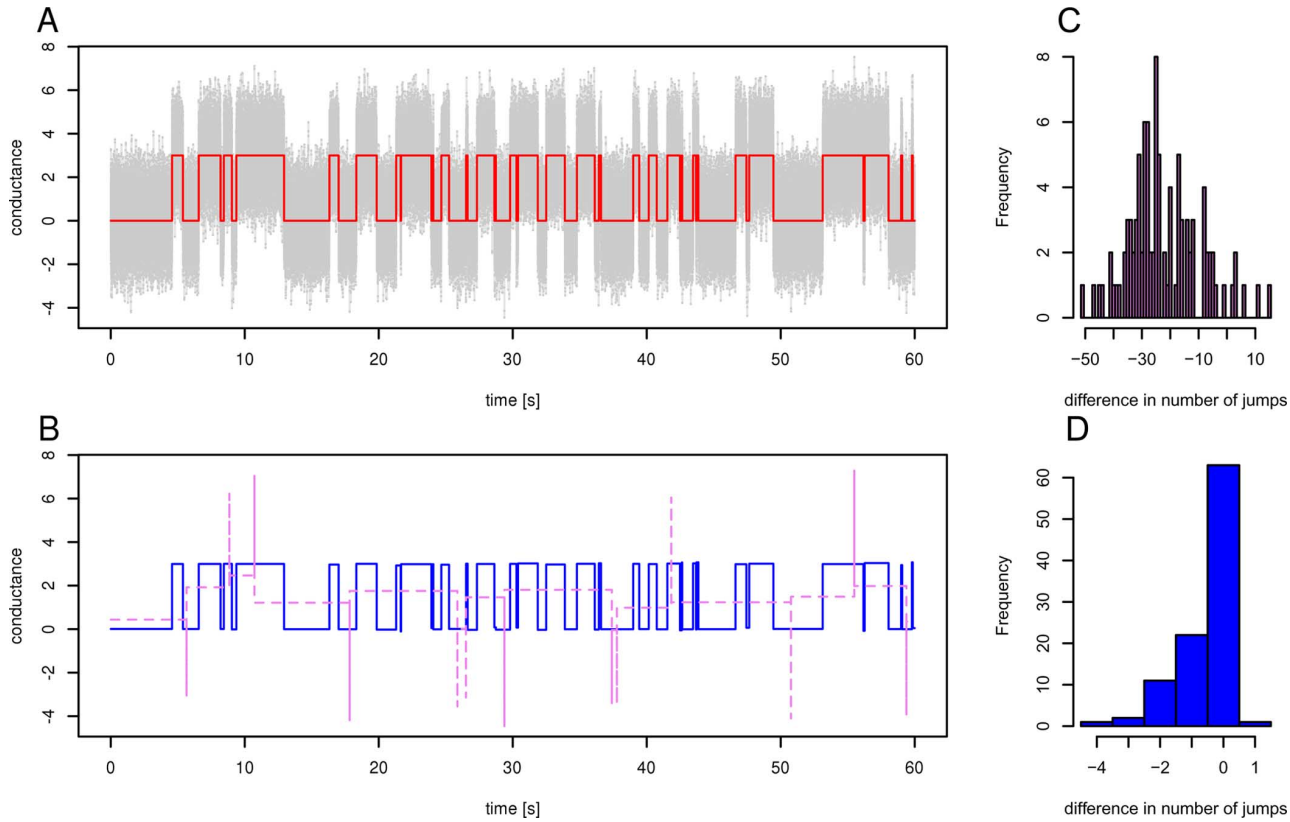

Fig. 1. (A) Simulated trace (grey background) of length 60 s, based on a two-state Markov chain with 1 Hz transition rates (solid red line) with noise added to give a SNR of 3 after filtering. (B) TRANSIT (dashed purple line) and J-SMURF (solid blue line) reconstructions. (C) Histogram of difference between number of detected and number of simulated jumps based on 100 simulated traces for TRANSIT. (D) Similar histogram for J-SMURF.

Munich, Germany). Briefly an electrically insulating membrane was generated by spreading a peptide containing GUV onto an aperture of a borosilicate glass chip. A DC potential (V_{m}) between -100 and 100 mV was applied. After observation of current transitions, data was recorded at a sampling rate of 10 kHz using a 1 kHz four-pole Bessel low-pass filter.

For synthetic and experimental details see Appendix A.

C. Analysis

For each trace, we first determined the noise's standard deviation using the robust estimator described above; these estimates were used as input both for J-SMURF and for TRANSIT [14]. For the experimental data, single electronic spikes were automatically detected and removed if necessary. No baseline correction was applied.

The SMRA was restricted to intervals of dyadic length for increased computational efficiency. The significance level was set to 5%; the corresponding 95 % quantiles were obtained by simulating 4,000 traces of pure, filtered noise for 60 s, leading to a critical value of 11.53. Where shorter traces are shown, these have been cut out of 60 s traces after the analysis; in particular, the critical values were not adapted a posteriori to the shorter time lengths.

Amplitude threshold was used for final idealization after determining the J-SMURF estimator for traces of experimental data in order to cope with small baseline drifts. These thresholds were chosen by the analyst based on the J-SMURF estimator.

The TRANSIT algorithm was reimplemented using only central differences, no forward differences, since all traces had been low-pass filtered.

The analyses were performed using the statistics software R [37], all algorithms having been implemented in a custom package called "stepR", available from http://stochastik.math.uni-goettingen.de/smuce; implementation details may be found in [29].

IV. Results

We validated J-SMURF on simulated as well as experimental data sets. While simulations enabled us to explore the potential of the multiresolution approach, analyses of experimental data sets showed its practical usability to idealize ion channel recordings. Gramicidin A provides an ideal test case for our approach, as it combines well-defined characteristics with a challengingly low conductivity [24]. As a proof of J-SMURF's general applicability we analyzed a novel fatty acid modified gA derivative to investigate the alteration of the channel's characteristics by changes in the amino acid sequence and lipid-peptide interactions [38].

A. Simulations

We simulated three scenarios using realistic parameters: first pure noise, then a two-state system with slow dynamics, finally one with fast dynamics; the two states corresponding to the simulated channel being open or closed. Each trace was simulated to be 60 s long; 100 traces were generated per simulation.

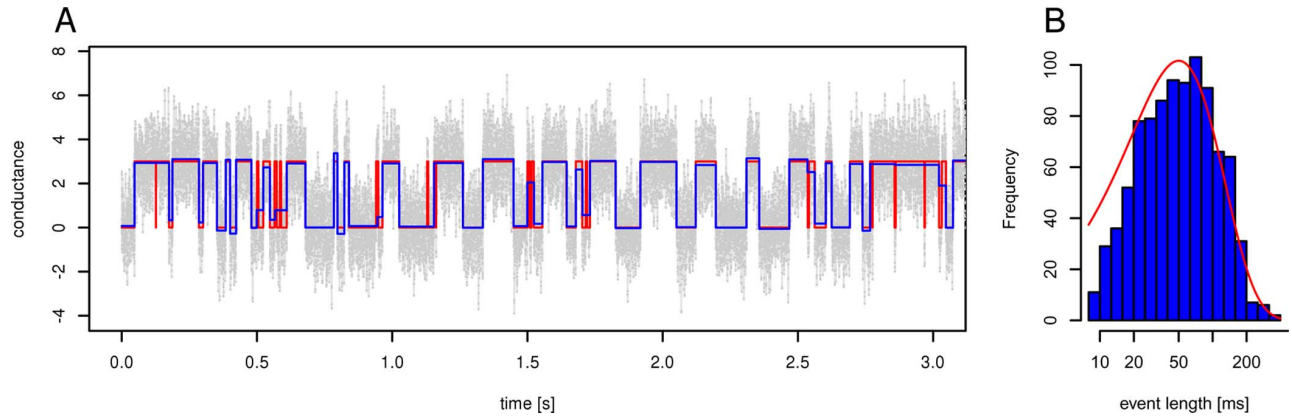

Fig. 2. (A) First 3 s of a simulated trace (grey background) of length 60 s, based on a two-state Markov chain with 20 Hz transition rates (solid red line) with noise added to give a SNR of 3 after filtering; J-SMURF reconstruction overlaid (solid blue line). (B) Histogram (blue) of length of events (log-scale) detected by J-SMURF based on the entire 60 s trace, and corresponding theoretical distribution (solid red line).

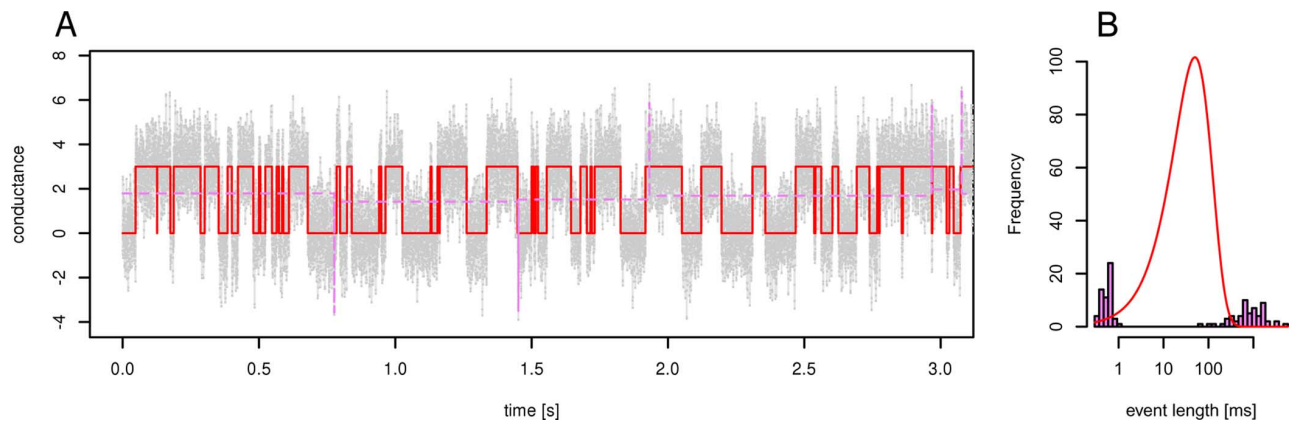

Fig. 3. (A) First 3 s of a simulated trace (grey background) of length 60 s, based on a two-state Markov chain with 20 Hz transition rates (solid red line) with noise added to give a SNR of 3 after filtering; TRANSIT reconstruction overlaid (dashed purple line). (B) Histogram (purple) of length of events (log-scale) detected by TRANSIT based on the entire 60 s trace, and corresponding theoretical distribution (solid red line).

1) Noise Only: First, we simply considered a constant signal with filtered noise added, hence no jumps should be detected. For the 100 simulated traces, TRANSIT detected between 34 and 104 jumps in a trace (61 on average). For the same traces, J-SMURF detected at most 1 jump per trace, which happened in 6 out of 100 traces. This is what was to be expected: for long time series, TRANSIT adds spurious jumps while J-SMURF obeys the global significance level of 5%, meaning that on average in only 5% of the traces a jump may erroneously get detected.

2) Slow Dynamics: Mimicking slow-dynamics ion channels, we simulated a two-state system with transition rates of 1 Hz, leading to an expected number of 60 jumps within the simulated 60 s time interval; the SNR was 3, i.e., the difference between the conductance levels was 3 while the filtered noise's standard deviation was set to 1. A simulated trace is shown in Fig. 1(A), with reconstructions by TRANSIT and J-SMURF shown in Fig. 1(B).

At this low SNR, TRANSIT no longer works reliably: in 100 simulated traces, it detected between 51 less and 15 jumps more than the true signal contained, see Fig. 1(C). Note that this shows that changing the threshold in either direction will not lead to reliable results: increasing the threshold will lead to even more spurious jumps being detected while decreasing it will result in even more being missed. In contrast, J-SMURF detected between 4 jumps too few and 1 jump too many, detecting the correct number of jumps in 63 traces, see Fig. 1(D).

3) Fast Dynamics: To explore the resolution limit of J-SMURF, we also simulated another two-state system but now with 20 Hz transition rates, see Fig. 2(A) for a sample trace. In 100 simulated traces, J-SMURF detected between 300 and 210 jumps less than truly simulated, 253 on average; this is to be compared with the $60\ \mathrm{s} \times 20\ \mathrm{Hz} = 1200$ expected events per trace. Fig. 2(B) visualizes the difference between the lengths of the events detected by J-SMURF and the theoretical distribution of event lengths. As expected, it is mostly the relatively short events (shorter than 20 ms) that are missed; however, the true rate can still be read off from the mode at around 50 ms. Note that this transition rate is fast in comparison to the simulated experimental conditions, namely the low SNR of 3 *after* applying a 1 kHz low-pass filter. Indeed, one may rescale time, e.g., by a factor of 20: if a 20 kHz filter was sufficient to obtain this SNR, about 1 ms long events could still be detected.

Fig. 3 shows the corresponding results of TRANSIT for these simulations. For the 100 simulated traces, TRANSIT detected between 982 and 1196 less jumps than actually were in the signal; the expected number of jumps was 1200.

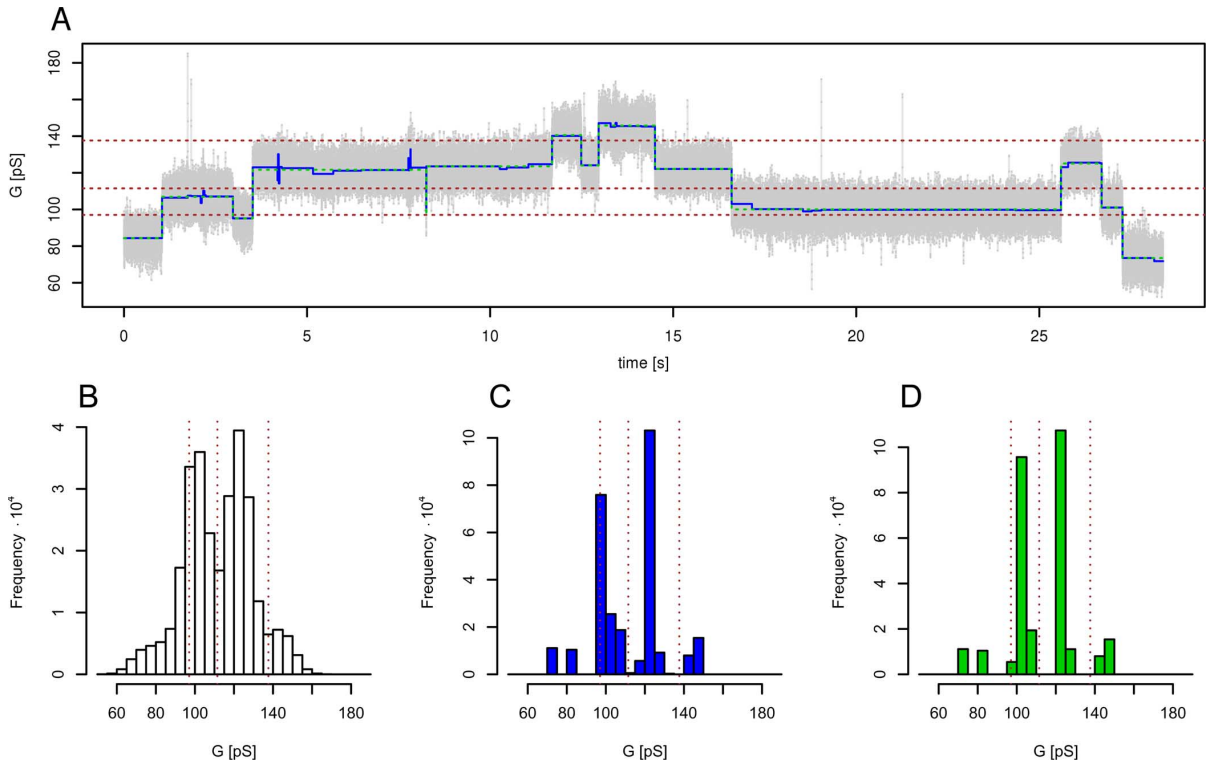

Fig. 4. (A) Time trace of conductance for gA, $V_m = -50$ mV, 28 s (grey background), with J-SMURF reconstruction (blue solid line) and idealization (green dashed line). (B) Point histogram of raw data. (C) Point histogram after applying J-SMURF. (D) Point histogram after idealization. Thresholds for idealization shown as dotted brown lines.

B. Experimental Data

1) Gramicidin A: Fig. 4(A) shows a characteristic conductance trace of gA with a typical SNR. Clearly, different conductance levels can be distinguished which are characteristic for the opening and closing of multiple channels. While J-SMURF reconstructed all visually recognizable events, it also included a certain amount of spurious jumps due to a drift in the baseline and background noise. Therefore, based on the J-SMURF estimator, Fig. 4(A), and the corresponding point histogram, Fig. 4(C), thresholds were determined by the analyst for a final idealization, Fig. 4(A), (D).

For some traces, fast gating of the channel could be observed, see Figs. 5 and 6. This phenomenon of gA channels is rarely observed and was previously attributed to the extent of the hydrophobic mismatch between the conducting dimer and the surrounding membrane, and variations in channel geometries at the bilayer/channel interface [39], [40]. The short conductance transitions are clearly too fast to be detected at all times given, as the rise time of the analog low-pass filter used is in the range of the event length, resulting in the conductance not to fully reach the corresponding level again. Nonetheless, many of the fast events could be detected while removing most of the noise where the signal remained constant. The additional fast gating results in a change of the evident channel dynamics. The short closing events and the diffusion triggered disassembly of a conducting channel are superimposed. Fortunately, both dynamics can be distinguished as shown in Fig. 6. Plotting conductance versus length of events shows many fast closing events while opening times are much longer, see Fig. 6(C).

2) Acylated gA Derivative: The analysis of a novel fatty gA derivative gives rise to the possibility to check for J-SMURF's capability to idealize recordings of an ion channel with largely unknown properties. The chemical structure of the derivative used is shown in Fig. 7(B). Attachment of fatty acids to the gA structure could largely influence the channels interaction with the surrounding lipids resulting in changes in channel dynamics and conductivity [41]–[44]. To achieve a fatty acid modification without disturbing the C-terminal ethanolamide moiety, the amino acid sequence was slightly modified. Fatty acid modifications are also known to influence the dynamic behavior of gA.

The C10 acylated gA shows typical opening and closing of the channel (Fig. 8). However, not obvious from the raw data, see Fig. 8(A)–(B), the derivative exhibits subgating, resulting in conductance transitions between two or three distinct sublevels while one, respectively two, channels are open, see Fig. 8(C). This is not visible from the raw data's point histogram in Fig. 8(B) due to the very low SNR associated with the subgating, showing the power of J-SMURF to detect an unexpected behavior in a conductance trace without any external input of the operator.

V. Discussion

J-SMURF clearly meets the criteria of an idealization method which we asked for: it is fully automatic, requiring input of the analyst only for the final idealization after the noise has been removed, and only in order to remove small artefacts caused by a baseline drift or similar effects on the background. It is universally applicable to data with low or high SNR, slow as

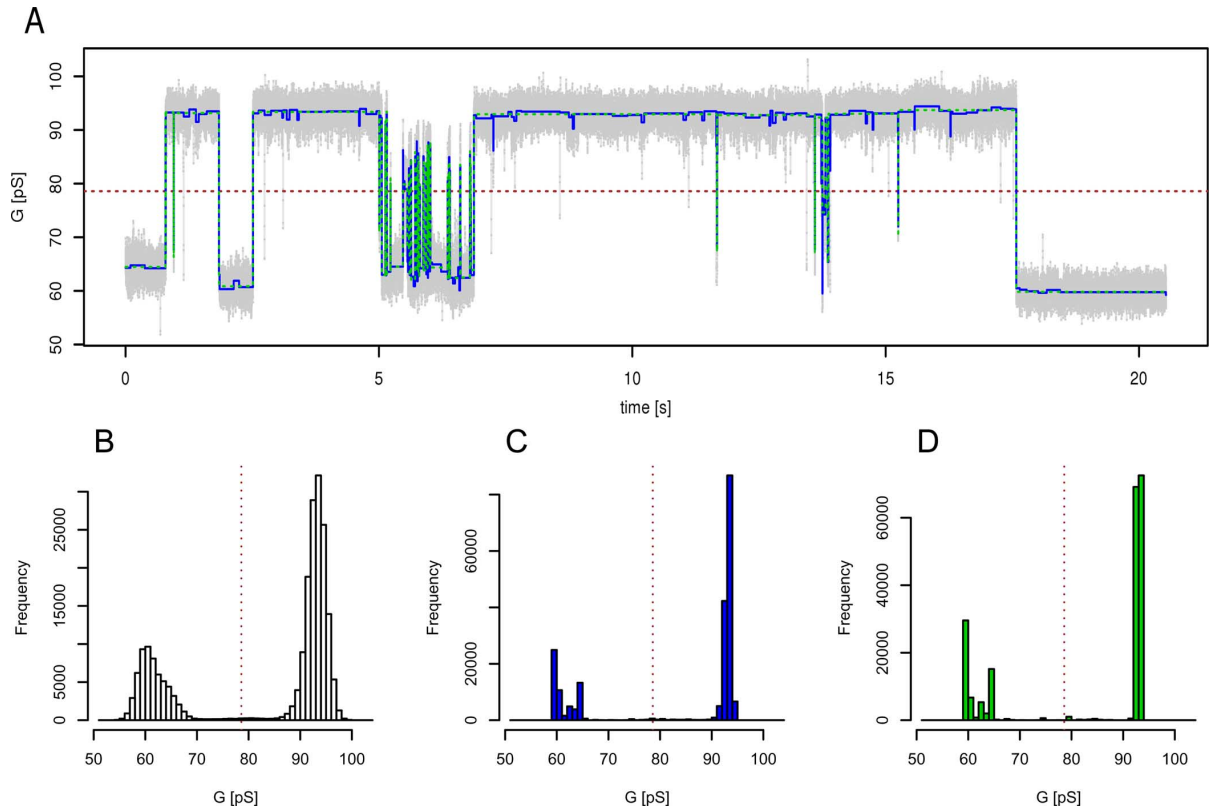

Fig. 5. (A) Time trace of conductance for gA, $V_{\mathrm{m}} = 100$ mV, 21 s (grey background), with J-SMURF reconstruction (blue solid line) and idealization (green dashed line). (B) Point histogram of raw data. (C) Point histogram after applying J-SMURF. (D) Point histogram after idealization. Thresholds for idealization shown as dotted brown lines.

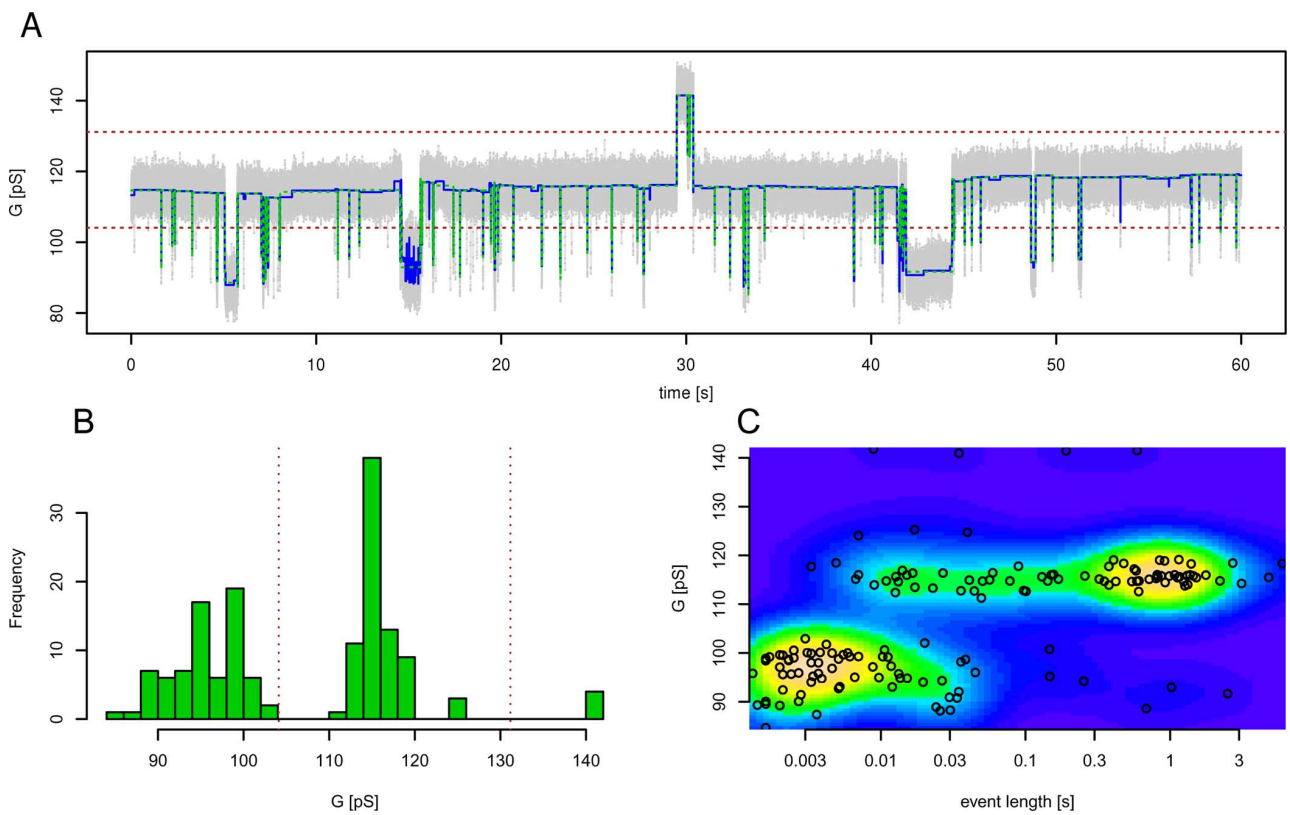

Fig. 6. (A) Time trace of conductance for gA, $V_{\mathrm{m}} = -100$ mV, 60 s (grey background), with J-SMURF reconstruction (blue solid line) and idealization (green dashed line). (B) Event histogram of idealization. (C) Idealization events' currents vs length (log-scale) shown as points, background obtained by smoothing. Thresholds for idealization shown as dotted brown lines.

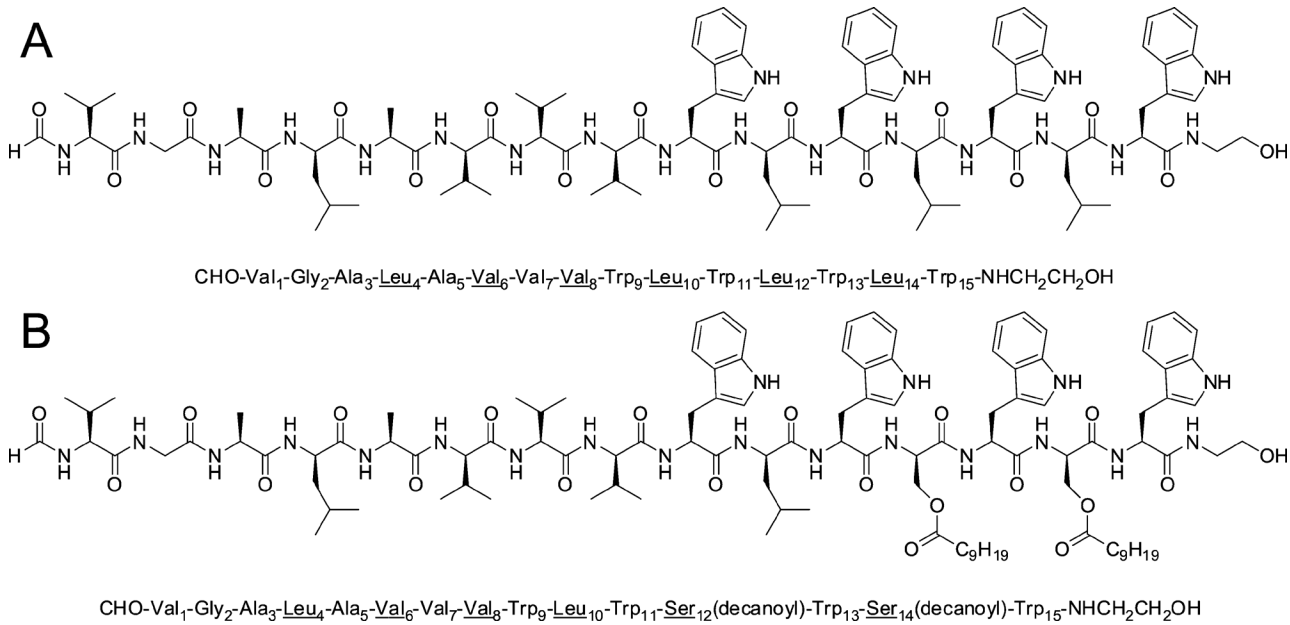

Fig. 7. (A) Chemical structure of [Val_1]gA corresponding to the amino acid sequence HCO-NH-L-Val_1-Gly_2-L-Ala_3-D-Leu_4-L-Ala_5-D-Val_6-L-Val_7-D-Val_8-L-Trp_9-D-Leu_{10}-L-Trp_{11}-D-Leu_{12}-L-Trp_{13}-D-Leu_{14}-L-Trp_{15}-CO-$NHCH_2CH_2OH$ (B) Correspoding structure of the gA derivative with D-Leu_{12} and D-Leu_{14} substituted by D-Ser(decanoyl).

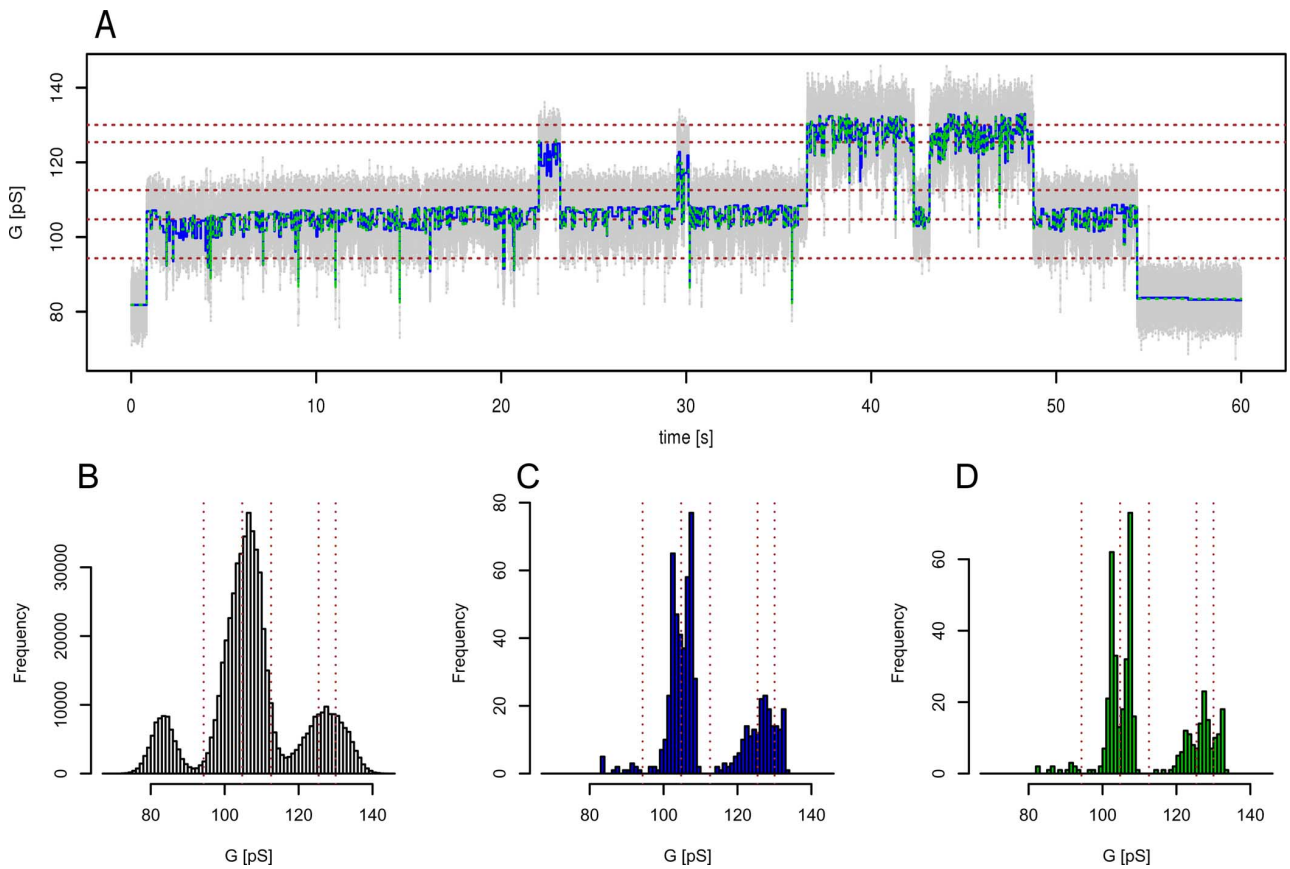

Fig. 8. (A) Time trace of conductance for the acylated gA derivative, $V_m = -50$ mV, 21 s (grey background), with J-SMURF reconstruction (blue solid line) and idealization (green dashed line). (B) Point histogram of raw data. (C) Event histogram after applying J-SMURF. (D) Event histogram after idealization. Thresholds for idealization shown as dotted brown lines.

well as fast dynamics; only if the events are very short in relation to the SNR and the analog low-pass filter that has been applied, J-SMURF is no longer able to detect them with statistical significance. Meanwhile, J-SMURF accounts for multiple testing, thus controlling the probability of erroneously detecting a jump.

In comparison, thresholding approaches like TRANSIT, when combined with an advance low-pass filter, inherently have a certain event length for which they are optimized, not making use of longer events if available. Although their filter may be adapted to some specific dynamics, this is difficult to achieve over the wider range of event lengths occurring, while at the same time requiring prior knowledge.

This was possible by adapting statistical multiresolution analyses for the situation where an analog filter has been applied before the analysis, rendering the noise colored. Hence, for J-SMURF, the only parameter that needs to be chosen is the significance level α which we took to be 5%, leading to the quantile q being 11.53. However, we empirically found the J-SMURF estimator to give qualitatively similar results when α was changed between 1% and 10%, say, with the corresponding quantiles between 12.17 and 11.19.

After idealization, one may already have a better understanding of the channel's behavior which one can then model appropriately in order to obtain parameter estimates, the latter being often based on the idealized data [5]. This typically has the advantage of being computationally more efficient: once the noise has been removed the parameters can be determined faster and more reliably e.g., by fitting a Markov model instead of a hidden Markov model, cf. [45]. While we have not discussed this issue here, it should be clear that such methods require reliable input data; in other words, the better the idealization, the better the parameter estimates obtained from it; analyzing the extent of this effect will be the goal of future research.

Indeed, there are still open questions to be addressed: if one wants to use the reconstruction obtained with J-SMURF for estimating the channel's dynamical parameters, e.g., employing a Markov model [5], one would need to correct for the very short events being missed, cf. [11], [12], [15]. Note that missing short events not only leads to an underestimation of their number but also to an overestimation of the events' lengths, cf. Fig. 2(B). However, we found that this apparently has little effect on the estimated conductances of longer durations, as only few, very short events are missed, cf. Fig. 2.

Algorithmically, J-SMURF is clearly more complex than a thresholding approach, resulting in computation times of several minutes on a standard desktop computer for the datasets considered here, see [29] for a discussion of its computational complexity. Improving J-SMURF's computational efficiency is therefore still an issue which we hope to address in the future.

Summarizing, the proposed combined jump segmentation and statistical multiresolution filter (J-SMURF) uses all available information, automatically adapting to the length of events, thus leading to an objective, universally applicable method for denoising and idealizing single ion channel recordings, where the detected events can be interpreted as being statistically significant.

APPENDIX A
FURTHER EXPERIMENTAL DETAILS

A. Gramicidin a Derivative Synthesis and Characterization

The synthesis of the gA derivative was performed via solid phase peptide synthesis (SPPS) using a peptide synthesizer (Microwave Peptide Synthesizer, CEM, Kamp-Lintfort, Germany, Applied Biosystems). A non-natural amino acid *N*-Fmoc-D-Ser(decanoyl)-OH was incorporated into the peptide sequence of gA by substituting the amino acids D-Leu_{12} and D-Leu_{14}. The C-terminal ethanolamide moiety of gA was introduced using a glycinol 2-chlorotrityl-resin. The building block *N*-Fmoc-D-Ser(decanoyl)-OH was synthesized by coupling *N*-Fmoc-D-Ser-OH with decanoyl chloride in TFA (trifluoracetic acid) [35]. Biomimetic *N*-terminal formylation was carried out on resin by using pentafluorophenyl formate in DCM [46]. Subsequently, the gA derivative was cleaved from the resin, the crude product was precipitated with cold diethyl ether (5 mL), purified by reverse phase high-performance liquid chromatography (HPLC) and characterized by high resolution mass spectrometry (HR-MS). Concentration of stock solutions of the freeze-dried peptide in 2,2,2-trifluoroethanol (TFE) were determined by UV-spectroscopy (Trp absorption at $= 283.9$ nm) [47].

Analytic Details:

HPLC (ReproSil100 ODS A C18, $(\mathrm{MeOH} + 0.1\%\ \mathrm{TFA})/(\mathrm{H_2O} + 0.1\%\ \mathrm{TFA})$ 9:1), $t_\mathrm{r} = 26.2$ min, HR-MS $C_{113}\mathrm{H}_{164}\mathrm{N}_{20}\mathrm{O}_{21}$ $[M + 2\mathrm{H}]^{2+}$ *calcd.:* 1069.51123, found: 1069.51132.

B. GUV Formation

To allow solvent-free recordings, gA and its derivative were introduced to the membrane during liposome formation. Giant unilamellar vesicles (GUVs) containing peptide were produced by the electroformation method [36]. A lipid solution of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine and cholesterol in a molar ratio of 9:1 (7.5 μL, $c = 10$ mM in $\mathrm{CHCl_3}$) was deposited on an indium tin oxide (ITO) coated cover slip and allowed to dry. Peptide solution was subsequently added to obtain a nominal lipid to peptide ratio of 10000:1 up to 1000:1. Residual solvent was removed under reduced pressure. Two ITO cover slips were electrically connected and the sealed chamber was filled with sorbitol solution (1 M) to rehydrate the film. GUV formation was carried out by applying a sinusoidal AC potential (3 V peak to peak, frequency of 5 Hz, 2 h, 20°C) using a voltage generator (33210A, Agilent Technologies, Böblingen, Germany) and was checked by optical microscopy.

C. Electrical Recordings

Electrical recordings were performed using the Port-a-Patch (Nanion Technologies GmbH, Munich, Germany). 6 μL of buffer solution (1 M KCl, 10 mM HEPES, pH 7.4, 20°C) were added to both sides of a borosilicate glass chip containing an aperture. A suspension of the peptide containing GUVs (2–6 μL) was added and spreading of a GUV was induced by appyling negative pressure (15–40 mbar). Successful formation of the solvent-free membrane was indicated by a resistance larger than 1 GΩ. The chip was rinsed (3×25 μL) with buffer and additional buffer was added (50 μL). A DC potential between -100 and 100 mV was applied. After observation of current transitions, data was recorded at a sampling rate of 10 kHz using an Axopatch 200B amplifier (Axon Instruments, Union City, CA). The signal was filtered with a four-pole Bessel low-pass filter of 1 kHz and digitized using an A/D converter (Digidata 1322, Axon Instruments). No further data modification was performed.

ACKNOWLEDGMENT

Generous support by the Deutsche Forschungsgemeinschaft (CRC 803, FOR 916) is gratefully acknowledged. O.M.S. thanks the IMPRS "Physics of Biological and Complex Systems" for financial support. The authors thank Erwin Neher, Eva K. Schmitt, and Ingo Mey for fruitful discussions.

REFERENCES

[1] B. Hille, *Ion Channels of Excitable Membranes*, 3rd ed. Sunderland, MA, USA: Sinauer, 2001.
[2] R. S. Kass, "The channelopathies: Novel insights into molecular and genetic mechanisms of human disease," *J. Clin. Invest.*, vol. 115, no. 8, pp. 1986–1989, 2005.
[3] J. P. Overington, B. Al-Lazikani, and A. L. Hopkins, "How many drug targets are there?," *Nat. Rev. Drug Discovery*, vol. 5, pp. 993–996, 2006.
[4] *Biological Membrane Ion Channels*, S. Chung, O. Andersen, and V. Krishnamurthy, Eds., 1st ed. New York: Springer, 2007.
[5] I. Siekmann, L. E. Wagner, II, D. Yule, C. Fox, D. Bryant, E. J. Crampin, and J. Sneyd, "MCMC estimation of markov models for ion channels," *Biophys. J.*, vol. 100, pp. 1919–1929, Apr. 2011.
[6] V. Krishnamurthy and G. G. Yin, "Controlled hidden markov models for dynamically adapting patch clamp experiments to estimate nernst potential of single-ion channels," *IEEE Trans. NanoBiosci.*, vol. 5, no. 2, pp. 115–125, 2006.
[7] F. G. Ball and J. A. Rice, "Stochastic models for ion channels: Introduction and bibliography," *Math. Biosci.*, vol. 112, pp. 189–206, 1992.
[8] C. Shelley, X. Niu, Y. Geng, and K. L. Magleby, "Coupling and cooperativity in voltage activation of a limited-state BK channel gating in saturating Ca^{2+}," *J. Gen. Physiol.*, vol. 135, no. 5, pp. 461–480, 2010.
[9] I. Goychuk, P. Hänggi, J. L. Vega, and S. Miret-Artés, "Non-Markovian stochastic resonance: Three-state model of ion channel gating," *Phys. Rev. E*, vol. 71, no. 6, pp. 061 906-1–061 906-11, 2005.
[10] S. Mercik and K. Weron, "Stochastic origins of the long-range correlations of ionic current fluctuations in membrane channels," *Phys. Rev. E*, vol. 63, no. 5, pp. 051 910-1–051 910-10, 2001.
[11] *Single-Channel Recording*, B. Sakmann and E. Neher, Eds., 2nd ed. New York: Plenum, 1995.
[12] D. Colquhoun, , N. B. Standen, P. T. A. Gray, and M. J. Whitaker, Eds., "Practical analysis of single channel records," in *Microelectrode Techiques. The Plymouth Workshop Handbook*. Cambridge, U.K.: Company of Biologists Ltd., 1987, pp. 83–104, .
[13] M. Basseville and A. Benveniste, "Design and comparative study of some sequential jump detection algorithms for digital signals," *IEEE Trans. Acoust.*, vol. 31, no. 3, pp. 521–535, Jun. 1983.
[14] A. M. J. VanDongen, "A new algorithm for idealizing single ion channel data containing multiple unknown conductance levels," *Biophys. J.*, vol. 70, pp. 1303–1315, Mar. 1996.
[15] A. L. Blatz and K. L. Magleby, "Correcting single channel data for missed events," *Biophys. J.*, vol. 49, pp. 967–980, May 1986.
[16] A. M. O'Connell, R. E. Koeppe, and O. S. Andersen, "Kinetics of gramicidin channel formation in lipid bilayers: Transmembrane monomer association," *Science*, vol. 250, no. 4985, pp. 1256–1259, 1990.
[17] O. S. Andersen, R. E. I. Koeppe, and B. Roux, "Gramicidin channels," *IEEE Trans. NanoBiosci.*, vol. 4, no. 1, pp. 10–20, Mar. 2005.
[18] W. R. Veatch, E. T. Fossel, and E. R. Blout, "Conformation of gramicidin A," *Biochemistry*, vol. 13, no. 26, pp. 5249–5256, 1974.
[19] R. R. Ketchem, B. Roux, and T. A. Cross, "High-resolution polypeptide structure in a lamellar phase lipid environment from solid state NMR derived orientational constraints," *Structure*, vol. 5, pp. 1655–1669, 1997.
[20] O. S. Andersen, H. J. Apell, E. Bamberg, D. D. Busath, R. E. Koeppe, F. J. Sigworth, G. Szabo, D. W. Urry, and A. Woolley, "Gramicidin channel controversy—The structure in a lipid environment," *Nat. Struct. Mol. Biol.*, vol. 6, p. 609, Jul. 1999.
[21] V. B. Myers and D. A. Haydon, "Ion transfer across lipid membranes in the presence of gramicidin A. II. the ion selectivity.," *Biochim. Biophys. Acta*, vol. 274, no. 2, pp. 313–322, 1972.
[22] J. A. Lundbæk, S. A. Collingwood, H. I. Ingólfsson, R. Kapoor, and O. Andersen, "Lipid bilayer regulation of membrane protein function: Gramicidin channels as molecular force probes," *J. R. Soc. Interface*, vol. 7, no. 44, pp. 373–395, 2010.
[23] J. Lundbæk, R. Koeppe, and O. Andersen, "Amphiphile regulation of ion channel function by changes in the bilayer spring constant," *Proc. Natl. Acad. Sci. USA*, vol. 107, no. 35, pp. 15 427–15 430, 2010.
[24] R. E. Koeppe and O. S. Anderson, "Engineering the gramicidin channel," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 25, no. 1, pp. 231–258, 1996.
[25] P. L. Davies and A. Kovac, "Local extremes, runs, strings and multiresolution," *Ann. Stat.*, vol. 29, no. 1, pp. 1–65, 2001.
[26] L. Dümbgen and V. G. Spokoiny, "Multiscale testing of qualitative hypotheses," *Ann. Stat.*, vol. 29, no. 1, pp. 124–152, 2001.
[27] D. Siegmund and B. Yakir, "Tail probabilities for the null distribution of scanning statistics," *Bernoulli*, vol. 6, no. 2, pp. 191–213, 2000.
[28] K. Frick, P. Marnitz, and A. Munk, "Statistical multiresolution dantzig estimation in imaging: Fundamental concepts and algorithmic framework," *Electron. J. Stat.*, vol. 6, pp. 231–268, 2012.
[29] K. Frick, A. Munk, and H. Sieling, "Multiscale change-point inference," ArXiv e-Prints no. 1301.7212, Jan. 2013, to appear in *J. Roy. Stat. Soc. B*, with discussion and rejoinder by the authors.
[30] C. Höhenrieder, "Nichtparametrische Volatilitäts- und Trendapproximation von Finanzdaten," 2008 [Online]. Available: http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=19097
[31] O. Wittich, A. Kempe, G. Winkler, and V. Liebscher, "Complexity penalized least squares estimators: Analytical results," *Math. Nachr.*, vol. 281, no. 4, pp. 582–595, 2008.
[32] L. Boysen, A. Kempe, A. Munk, V. Liebscher, and O. Wittich, "Consistencies and rates of convergence of jump penalized least squares estimators," *Ann. Stat.*, vol. 37, pp. 157–183, 2009.
[33] F. Friedrich, A. Kempe, V. Liebscher, and G. Winkler, "Complexity penalized M-estimation: Fast computation," *J. Comput. Graph. Stat.*, vol. 17, no. 1, pp. 201–224, 2008.
[34] H. Dette, A. Munk, and T. Wagner, "Estimating the variance in nonparametric regression—What is a reasonable choice?," *J. R. Stat. Soc. B, Stat. Methodol.*, vol. 60, no. 4, pp. 751–764, 1998.
[35] G. Portella, T. Polupanow, F. Zocher, D. A. Boytsov, P. Pohl, U. Diederichsen, and B. L. de Groot, "Design of peptide-membrane interactions to modulate single-file water transport through modified Gramicidin channels," *Biophys. J.*, vol. 103, no. 8, pp. 1698–1705, 2012.
[36] M. I. Angelova and D. S. Dimitrov, "Liposome electroformation," *Faraday Discuss. Chem. Soc.*, vol. 81, pp. 303–311, 1986.
[37] R Development Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing. Vienna, Austria, 2012 [Online]. Available: http://www.R-project.org/
[38] D. A. Kelkar and A. Chattopadhyay, "The gramicidin ion channel: A model membrane protein," *Biochim. Biophys. Acta, Biomembr.*, vol. 1768, no. 9, pp. 2011–2025, 2007.
[39] A. Ring, "Brief closures of gramicidin A channels in lipid bilayer membranes," *Biochim. Biophys. Acta, Biomembr.*, vol. 856, pp. 646–653, 1986.
[40] K. M. Armstrong and S. Cukierman, "On the origin of closing flickers in gramicidin channels: a new hypothesis," *Biophys. J.*, vol. 82, pp. 1329–1337, 2002.
[41] R. E. Koeppe, J. A. Paczkowski, and W. L. Whaley, "Gramicidin K, a new linear channel-forming gramicidin from bacillus brevis," *Biochemistry*, vol. 24, pp. 2822–2826, Jun. 1985.
[42] T. Vogt, J. A. Killian, B. De Kruijff, and O. Andersen, "Influence of acylation on the channel characteristics of gramicidin A," *Biochemistry*, vol. 31, pp. 7320–7324, 1992.
[43] T. C. Vogt, J. A. Killian, R. A. Demel, and B. De Kruijff, "Synthesis of acylated gramicidins and the influence of acylation on the interfacial properties and conformational behavior of gramicidin A," *Biochim. Biophys. Acta, Biomembr.*, vol. 1069, pp. 157–164, Nov. 1991.
[44] A. Benayad, D. Benamar, N. Van Mau, G. Page, and F. Heitz, "Single channel and monolayer studies of acylated gramicidin A: Influence of the length of the alkyl group," *Eur. Biophys. J.*, vol. 20, pp. 209–213, Jan. 1991.
[45] F. Qin, "Restoration of single-channel currents using the segmental k-means method based on hidden Markov modeling," *Biophys. J.*, vol. 86, pp. 1488–1501, Mar. 2004.
[46] L. Kisfaludy and L. Oetvoes, Jr, "Rapid and selective formylation with pentafluorophenyl formate," *Synthesis*, vol. 1987, no. 5, pp. 510–510, 1987.
[47] H. Edelhoch, "Spectroscopic determination of tryptophan and tyrosine in proteins," *Biochemistry*, vol. 6, no. 7, pp. 1948–1954, 1967.