Adjustments for the Display of Quantized Ion Channel Dwell Times in Histograms with Logarithmic Bins

J. Alex Stark and Stephen B. Hladky

National Institute of Statistical Sciences, Research Triangle Park, North Carolina 27709-4006 USA, and Department of Pharmacology, University of Cambridge, Cambridge CB2 1QJ, England

ABSTRACT Dwell-time histograms are often plotted as part of patch-clamp investigations of ion channel currents. The advantages of plotting these histograms with a logarithmic time axis were demonstrated by Blatz and Magleby (1986, *J. Physiol. (Lond.).* 378:141–174), McManus et al. (1987, *Pflügers Arch.* 410:530–553), and Sigworth and Sine (1987, *Biophys. J.* 52:1047–1054). Sigworth and Sine argued that the interpretation of such histograms is simplified if the counts are presented in a manner similar to that of a probability density function. However, when ion channel records are recorded as a discrete time series, the dwell times are quantized. As a result, the mapping of dwell times to logarithmically spaced bins is highly irregular; bins may be empty, and significant irregularities may extend beyond the duration of 100 samples. Using simple approximations based on the nature of the binning process and the transformation rules for probability density functions, we develop adjustments for the display of the counts to compensate for this effect. Tests with simulated data suggest that this procedure provides a faithful representation of the data.

INTRODUCTION

The dwell-time histogram is one of the most common forms of presentation for the results of patch-clamp investigations of ion channel currents. Blatz and Magleby (1986), McManus et al. (1987), and Sigworth and Sine (1987) have argued persuasively that it is most useful to plot such histograms over the logarithms of the dwell times; this makes it possible to clearly present in a single plot dwell times spanning several orders of magnitude and makes it easier to obtain a visual impression of the number of exponential components in the dwell-time distribution. However, the procedures for constructing these histograms are not entirely straightforward, because the recorded data are quantized as multiples of the sampling interval. When linearly quantized data are presented logarithmically the number of different, discrete dwell times mapped to each logarithmic bin varies. Log bins corresponding to the shortest dwell times receive the recorded occurrences or counts for only a few different dwell times or even for none, while log bins corresponding to long dwell times receive the counts for many different dwell times. To account for these variations McManus et al. (1987) chose to plot the ratio of the number of counts to the number of sample intervals mapped to that bin (i.e., the average counts per sample interval). Because the counts per sample interval become very small for long dwell times, they used a log-log presentation. For an exponential distribution of dwell times this gives the appearance of a plateau with a steep roll-off toward longer dwell times starting at about the mean. Sigworth and Sine (1987), who started from interpolated data in which dwell times are continuously distributed, argued that it is clearer to present the counts rather than the counts per sample interval. Attractive features of this style of presentation include the following:

- The rule for transforming the mean counts per linear bin to mean counts per log bin is the same as for the transformation of a probability density function (pdf).
- The spread of an exponential distribution plotted on a linear time axis increases with its mean, whereas when transformed and plotted on a log axis, such a distribution has invariant shape and spread; the density is simply translated along the time axis.
- When the distribution is transformed and plotted on a log axis the relative weights of well-separated exponential components can be estimated from their heights, and the mean dwell times can be estimated from the locations of the peaks.

We show here how to construct this type of plot when the dwell times are discrete instead of continuously distributed.

An example of the difficulties encountered when the dwell times are multiples of the sample interval is introduced in the next section. There is a markedly irregular variation in the number of counts in neighboring log bins; some bins remain empty and significant variations in counts extend to dwell times of 100 samples (see Fig. 1). McManus et al. (1987) dealt with these irregularities in their log-log presentation by constructing an empirical decoding file listing the midpoint time and number of mapped dwell times for each log bin. In the third section we build on the ideas they presented, introducing an automatic procedure for characterizing the binning process and extending the correction of their Eq. 24 to a pdf style of plot. We have chosen

Received for publication 24 May 1999 and in final form 30 September 1999.

Address reprint requests to Dr. J. Alex Stark, National Institute of Statistical Sciences, Research Triangle Park, NC 27709-4006, USA. E-mail: stark@niss.org.

Matlab code for adjusting logarithmic histograms and for the examples presented in this paper is available at http://www.phar.com.ac.uk/RI/sbh/logadjst.zip.

0006-3495/00/02/662/06 $2.00

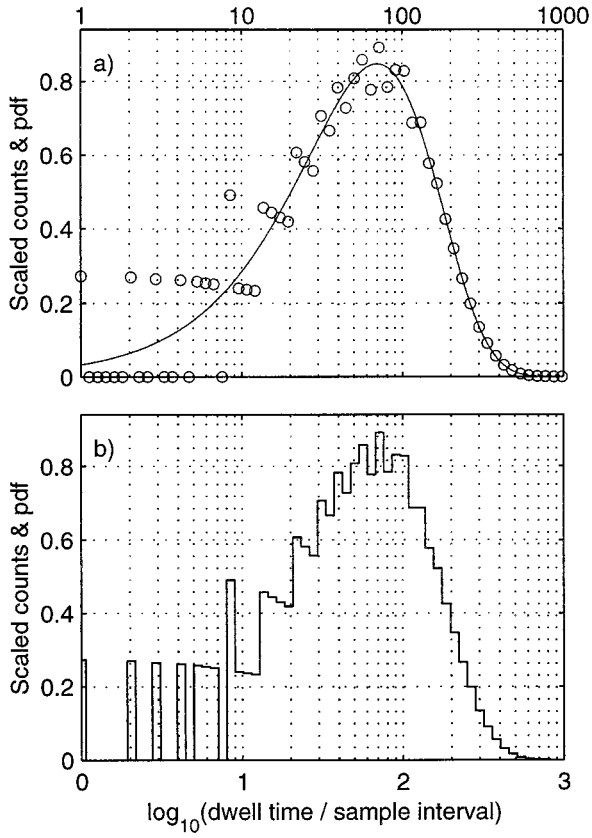

FIGURE 1 The observed counts in logarithmic bins of the rounded values from a single exponential distribution of dwell times with mean 70. The dwell times were mapped into bin numbers, using the formula in Eq. 1. The counts are plotted scaled by ω/N, where $1/\omega$ is the width of a log bin and N is the total number of counts. In part *a* the scaled counts are plotted as points with the equivalent transformed exponential superimposed as a pdf, and in *b* the scaled counts are plotted in the more familiar form of a traditional histogram. Significant irregularities in the binning extend to bins 39 and 40, to which dwell times around 103 and 116 are mapped. There are also errors in the abscissae, especially for small dwell times: the bin to which dwell time 10 has been mapped is centered between the grid lines for times 9 and 10. Note that the value of the scaled counts is zero in a number of bins corresponding to small dwell times, i.e., no counts are mapped to these bins.

to plot the estimated pdf as the ordinate so that areas on the transformed plot can be interpreted as probabilities. The square-root transformation suggested by Sigworth and Sine (1987) can easily be applied as a subsequent step if desired. The example is synthetic, but we have found that applying the same corrections to real results yields similar improvements.

This paper is primarily concerned with the visual presentation of logarithmically binned histograms. McManus et al. (1987) consider in detail the errors introduced into subsequent analysis by logarithmic binning. While simulations suggest that the procedures developed here are visually accurate, they are still approximations, and it will be more accurate if the original rather than the transformed data are used for numerical fitting (as in Qin et al., 1996, for example).

CHARACTERIZING THE BINNING PROCESS

Quantized logarithmic bins

When logarithmic binning is used for discrete dwell times, the range of dwell times that map to each bin is very irregular, even for longer values. Throughout this article we will follow an example in which dwell times τ, expressed as multiples of the sample interval, are mapped to bin numbers x by scaling their base 10 logarithms and rounding. This can be represented by the formula

$$b(\tau) = \langle \omega \log_{10}(\tau) \rangle, \quad (1)$$

where $\langle\ \rangle$ denotes rounding to the nearest integer and ω gives the number of bins per decade; we have chosen 19.385 because this highlights some problems more clearly than an integer value.

An example histogram was produced by generating 30,000 points from an exponential distribution with mean 70. The points were chosen deterministically; values of $\exp(-\tau/70)$ were spaced uniformly (see Press et al., 1992, Section 7.2). The counts generated using the mapping of Eq. 1 are displayed in Fig. 1, along with the theoretical distribution $g(z)$, given by

$$g(z) = \ln(10) \exp_{10}\left[z - z_m - \frac{\exp_{10}(z - z_m)}{\ln(10)}\right]. \quad (2)$$

where $z = \log_{10}(\tau)$ and $z_m = \log_{10}(70)$. The counts are scaled to correspond to a pdf when the abscissae are interpreted as base 10 logarithms. Significant irregularities in the counts extend to high bin numbers. In this case the count for times 97–109 (bin 39) is too high, whereas that for times 110–122 (bin 40) is too low. In addition to the variation in the number of times mapping to each bin, the abscissae can be misleading. In this example the count for time 10, which maps to bin 19, is placed between the scaled logs of times 9 and 10. The results generated with alternative forms of rounding, such as up or down, are generally even worse; these can arise implicitly when data are rebinned. The correction factors developed below can account for any rounding scheme.

Dwell-time ranges for each bin

Table 1, which lists the destination bin numbers given by Eq. 1 for a range of times, illustrates the discontinuous variation in the number of quantized durations that map to each bin. Bins 1–5 will remain empty, and while bin 18 will collect the counts for durations 8 and 9, only that for 10 maps to bin 19.

Characterizing the bin mapping is awkward, especially because numerical effects make a simple inversion of $b(\tau)$ difficult. There are two sources of error: mismatches between the software that gathers the histogram data and that which processes it, and numerical artefacts within one pro-

TABLE 1 The mapping $b(\tau)$ from quantized dwell time τ to logarithmic bins, x

τ	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9
00+		0	6	9	12	14	15	16	18	18
10+	19	20	21	22	22	23	23	24	24	25
20+	25	26	26	26	27	27	27	28	28	28
30+	29	29	29	29	30	30	30	30	31	31

Base 10 logarithms of the dwell times were scaled by 19.385 and rounded to the nearest integer, as in Eq. 1. For τ given as the sum of a value in the top row and a value in the leftmost column, the index of the log bin, x, is given at the intersection of the corresponding row and column.

gram. An example of the latter is that, in tests, we found it necessary to account for the errors in the computed value of exp(ln(10)). These can be addressed through the use of a decoding file, as suggested in Section I.4(b) of McManus et al. (1987). We have found it most convenient not to attempt to invert $b(\tau)$ directly but rather to summarize the mapping information in a function, $a(x)$, which provides the lowest dwell time that maps to bin number x or higher. An example is shown in Table 2. This could be calculated exactly by lookup from values of $b(\tau)$ in an extended version of Table 1. Alternatively one can employ a generating function and test values on either side. In these tests (with double-precision floating-point arithmetic),

$$a(x) = \lceil 10^{(x-1/2)/\omega} - \eta \rceil, \tag{3}$$

where $\lceil\ \rceil$ denotes rounding up, produced correct results with $\eta = 10^{-10}$. The factor η ensures that values that should be at or just below integers are not erroneously rounded up.

The lower bound on the range of quantized dwell times that map to a bin x is, by definition, $a(x)$. The shortest dwell time to map to a higher bin is $a(x + 1)$, even if bin $x + 1$ is unmapped. Hence the number of dwell times that map to bin x is

$$s(x) = a(x + 1) - a(x). \tag{4}$$

Example values are set out in Table 3.

TABLE 2 The shortest dwell time, $a(x)$, which is mapped to bin x or higher

x	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9
00+	1	2	2	2	2	2	2	3	3	3
10+	4	4	4	5	5	6	7	8	8	10
20+	11	12	13	15	17	19	21	24	27	30
30+	34	38	43	48	54	61	68	77	86	97

For x given by the sum of a value in the top row and a value in the leftmost column, the intersection of the corresponding row and column contains the minimum value of τ for which $b(\tau)$ as set out in Table 1 is greater than or equal to x.

TABLE 3 The number of quantized dwell times, $s(x)$, in the range that maps to each bin x

x	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9
00+	1	0	0	0	0	0	1	0	0	1
10+	0	0	1	0	1	1	1	0	2	1
20+	1	1	2	2	2	2	3	3	3	4
30+	4	5	5	6	7	7	9	9	11	13

This equals $a(x + 1) - a(x)$, the discrete difference of the sequence in Table 2. For a value of x given as the sum of the values in the leftmost column and the top row, $s(x)$ is given at the intersection of the corresponding row and column.

Rebinning

There are situations in which it is possible to collect counts with a higher bin density than would be used for plotting. Bins would then be amalgamated to simplify plots and reduce statistical variation in the counts per bin. In such cases $b(\tau)$ then represents the combination of the original binning and the rebinning processes. Note, however, that simple uniform amalgamation of neighboring bins has a *rounding down* effect (when indexed from 0 as here). Rounding to the nearest integer can be preserved if the bin density is reduced by an odd factor and the original histogram is padded with empty bins at its ends.

ADJUSTMENTS

Abscissa adjustment

Consider the task of displaying the observations in a bin by a point. The scaling and rounding operations affect not only the ordinates, but also the abscissae. The simplest case is where exactly one dwell time maps to a bin. Plotting the point at the midpoint of the logarithmic bin does not correctly represent the logarithm of the dwell time that maps to it because of the rounding operation; e.g., in this example without adjustment, dwell time 10, which maps to bin 19, is displayed between the logarithms of 9 and 10. In this case it makes sense to remap the value of x to the appropriate scaled logarithm, $\hat{z}$:

$$\hat{z} = \log_{10}(a(x)), \tag{5}$$

which correctly maps dwell time 10 to 1 in the example. This can be readily generalized for bins that receive counts from more than one dwell time. A point representation of a bin count is most meaningful if it is located centrally in the range of dwell times that map to it, i.e., using a relation such as

$$\hat{z} = (1/2)\,[\log_{10}(a(x)) + \log_{10}(a(x + 1) - 1)]. \tag{6}$$

For example, using this rule, the adjusted counts for dwell times 8 and 9, which both map to bin 18, will be displayed at $(1/2)\log_{10}(8 \times 9)$. Any available knowledge about the quantization process should be taken into account. If the

continuous dwell times are known to be rounded to the nearest integer, as in this example, then the range being mapped into bin x starts at $a(x) - 1/2$ and ends at $a(x + 1) - 1/2$.

Ordinate adjustment for points

In this section we will combine the use of Eq. 6 with an adjustment to the ordinate to produce a graph of the form shown in Fig. 2. We begin by considering the case of a discrete display. The objective is to estimate the pdf for the logarithmic scale from the observed counts for the dwell times.

The impression conveyed by a graph for the probability of an event being observed with any dwell time range should be independent of the choice of axes used. The probability of an event having a duration between τ_1 and τ_2 can be expressed in terms of pdfs $f(\tau)$ and $g(z)$ for linear and logarithmic scales, respectively:

$$\int_{\tau_1}^{\tau_2} f(\tau)\, \mathrm{d}t = \int_{\log_{10}(\tau_1)}^{\log_{10}(\tau_2)} g(z)\, \mathrm{d}z. \tag{7}$$

For small-duration ranges one can accurately approximate the pdfs as linear. Thus for z and τ in the middle of the range,

$$g(z) \approx \frac{f(\tau)\,(\tau_2 - \tau_1)}{\log_{10}(\tau_2) - \log_{10}(\tau_1)}. \tag{8}$$

An estimate of $f(\tau)$ can be obtained by dividing the fraction of counts falling in bin x by $s(x)$ (McManus et al., 1987, Eq. 24). Hence, the adjustment factor that should be applied to the counts to obtain an estimate of $g(z)$ is

$$\psi(x) = \frac{(\tau_2 - \tau_1)}{N\, s(x)[\log_{10}(\tau_2) - \log_{10}(\tau_1)]}. \tag{9}$$

where N is the total number of counts. If the sparsely separated lower bins are deemed to be interpreted in a fashion similar to that of the higher ones, which are virtually uniformly spaced, then we should use a range of half a bin, logarithmically, on either side. In other words,

$$\log_{10}(\tau_1) = \hat{z} - \frac{1}{2\omega}, \tag{10}$$

$$\log_{10}(\tau_2) = \hat{z} + \frac{1}{2\omega}. \tag{11}$$

This is the scheme used to generate Fig. 2.

Traditional histogram display

The adjustments introduced in the preceding section produce an estimate of the probability density function over $z = \log_{10}(t)$. It is also possible to generate a histogram in the traditional style—a connected staircase as in Fig. 3. We know that the upper quantized dwell time mapped to a bin is $a(x + 1) - 1$ and that the lower time for the next bin is $a(x + 1)$. Choosing arbitrarily to draw boundaries half-way between these discrete linear dwell times and thus drawing

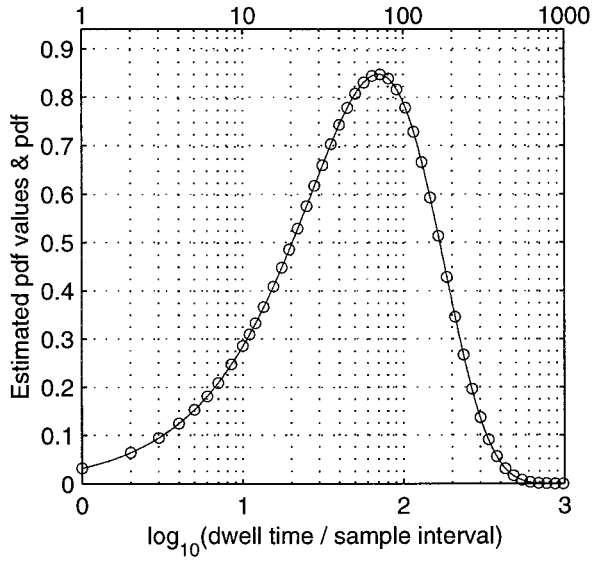

FIGURE 2 A plot of the adjusted data values as estimates of the pdf. The theoretical curve is the same as in Fig. 1. The estimated pdf values are the counts multiplied by $\psi(x)$, the adjustment factor defined in Eq. 9, using log bin boundaries defined in Eqs. 10 and 11.

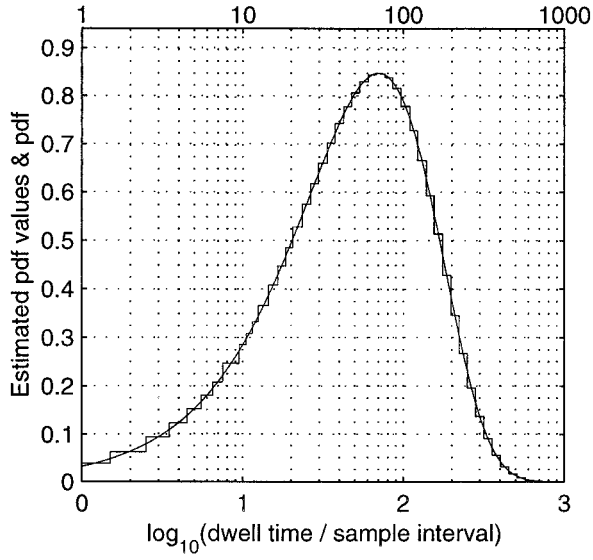

FIGURE 3 A traditional style of histogram plot. For each bin the range of quantized dwell times was $a(x)$ to $a(x+1) - 1$. The boundaries between bins are the midpoints in linear time between endpoints of neighboring ranges; bins that were unmapped were discarded. The number of counts in each bin is proportional to the area. In this case $\psi(x)$ was calculated using Eq. 9 with Eqs. 12 and 13.

a bin extending from a lower limit,

$$\tau_1 = a(x) - 1/2, \tag{12}$$

to an upper limit,

$$\tau_2 = a(x + 1) - 1/2, \tag{13}$$

the number of dwell times mapped to the bin is

$$s(x) = a(x + 1) - a(x) = \tau_2 - \tau_1. \tag{14}$$

The adjustment factor in Eq. 9 becomes

$$\psi(x) = \frac{1}{N[\log_{10}(\tau_2) - \log_{10}(\tau_1)]}, \tag{15}$$

and the area of the bin is simply proportional to the sum of the counts for the corresponding discrete dwell times. McManus et al. (1987) discuss in detail how these binned counts can be used in subsequent analysis. It should be noted that while the bin boundaries do correspond to the correct discrete dwell times, the exact positioning between these dwell times has been chosen arbitrarily, and thus the boundaries do not indicate the precise range of continuous dwell times mapped to each bin. Largely to avoid conveying a false impression, we prefer the pdf style of plot, as in Fig. 2, for the visual presentation of data.

Bounds on the correction factor

It is possible to derive an expression that bounds the deviation from unity of the ratio of the observed and adjusted count values:

$$\left|\frac{1}{\psi(\omega \log_{10}(\tau))} - 1\right| \lessapprox \frac{1}{\tau(10^{1/2\omega} - 10^{-1/2\omega})}. \tag{16}$$

As can be verified empirically, values occur close to this bound, and hence substantial deviations occur for durations longer than 100 samples.

DISCUSSION

The uses and advantages of using a logarithmic time axis for the display of dwell-time data have been set out by by Blatz and Magleby (1986), Sigworth and Sine (1987), and McManus et al. (1987). The principal impediment arises because the experimentally observed dwell times are necessarily multiples of the sample interval. The irregularities in histograms generated from the logarithms of discrete dwell times can be severe for short dwell times and can even be significant for durations as long as 100 samples. Furthermore, the midpoints of the logarithmic bins that display the data for small dwell times do not occur at the log of the duration they represent. The irregularities can be alleviated by interpolation of the experimental record between data points (see Sigworth and Sine, 1987) or by maintaining a table that records for each bin "the actual midtime of the bin and bin width" for use in decoding the histogram (McManus et al., 1987). The method outlined in this paper allows the displayed values to be adjusted so as to remove the irregularities.

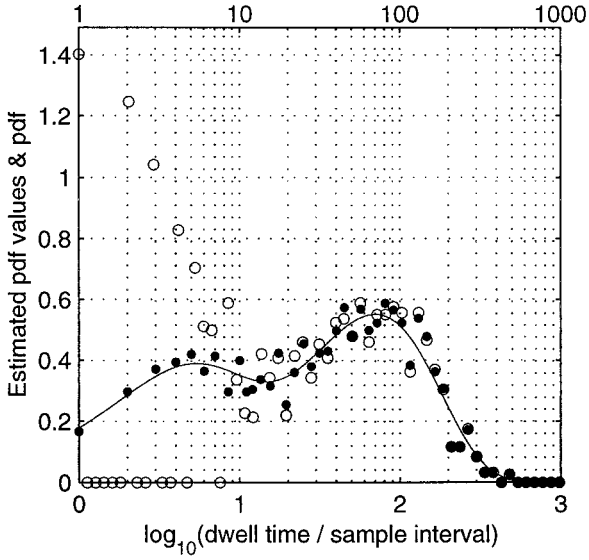

FIGURE 4 The observed counts (○) and estimated pdf (●) for 3000 dwell times τ drawn at random from the pdf $(0.35/4)\exp(-\tau/4) + (0.65/70)\exp(-\tau/70)$, plotted as a line after transformation.

The adjustment factors are based on simple approximations, the description of the binning process, and rules for the transformation of pdfs. These allow calculation of point estimates of the distribution for the logarithmic scale, based directly on the frequency of observation of the discrete dwell times. We recommend this method of display. However, if preferred it is possible to present histograms in the more traditional style of a connected staircase in which the bin boundaries indicate the range of the *discrete* dwell times mapped to the bin, the height is an estimate of the probability density function, and the area is proportional to the number of counts in the bin.

Fig. 2 shows that the adjustments proposed are able to recover, essentially without error, a single exponential distribution of dwell times. A more realistic example is provided in Fig. 4, which plots dwell times drawn at random from a double-exponential distribution. For logarithmic binning defined by Eq. 1, the relatively simple adjustments provided by Eqs. 6 and 9 with Eqs. 10 and 11 provide a visually accurate representation of the data.

This research was supported in part by grant 8/E03204 from the Biotechnology and Biological Sciences Research Council, and one author (JAS) was supported in part by Christ's College, Cambridge.

REFERENCES

Blatz, A. L., and K. L. Magleby. 1986. Quantitative description of three modes of activity of fast chloride channels from rat skeletal muscle. *J. Physiol. (Lond.).* 378:141–174.

McManus, O. B., A. L. Blatz, and K. L. Magleby. 1987. Sampling, log binning, fitting and plotting durations of open and shut intervals from single channels and the effects of noise. *Pflügers Arch. Eur. J. Physiol.* 410:530–553.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical Recipes in C: The Art of Scientific Computing, 2nd Ed. Cambridge University Press, Cambridge.

Qin, F., A. Auerbach, and F. Sachs. 1996. Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* 70:264–280.

Sigworth, F. J., and S. M. Sine. 1987. Data transformations for improved display and fitting of single-channel dwell time histograms. *Biophys. J.* 52:1047–1054.