A Multiple-Technique Approach for the Reproducible Inference of Bio-electrochemical Reaction Parameters

Henry O. Lloyd-Laney,[†] Nick D. J. Yates,[‡] Martin J. Robinson,[†] Alice R. Hewson,[‡] Jess Branch,[¶] Glyn R. Hemsworth,[¶] Alan M. Bond,[§] Alison Parkin,[*,‡] and David J. Gavaghan[*,†]

[†]*Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford, OX1 3QD United Kingdom*

[‡]*Department of Chemistry, University of York, Heslington, York, YO10 5DD, United Kingdom*

[¶]*Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom*

[§]*School of Chemistry, Monash University, Clayton, Vic., 3800 Australia*

E-mail: alison.parkin@york.ac.uk; david.gavaghan@dtc.ox.ac.uk

Abstract

Uncovering the secrets of the biological Faradaic reactions, essential to the understanding of complex metalloenzymes, requires an information recovery process that is robust, rapid and replicable. This paper is a description of the workflow that was developed over the course of inferring chemical reaction parameters for a simple protein system, a bacterial cytochrome domain from *Cellvibrio japonicus*. This was a challenging task, as the signal-to-noise ratio in such protein-film voltammetry experiments is significantly lowered relative to the voltammetric data generated by low weight molecular compounds. These challenges were overcome by using a multiple-technique approach, which compensates for the difficulties inherent to analysis of the individual voltammetry experiments. The parameters inferred are robust across multiple experiments performed for different preparations of the protein. This is an important proof-of-concept result for analysis of more complex metalloenzymes, which incorporate catalytic processes and multiple internal electron-transfer sites.

Introduction

Of the myriad chemistries that are required for the existence of life, some of the most fundamental such as the splitting of water, the fixation of nitrogen, and carbon capture are underpinned by electron-transfer reactions. Such catalytic redox processes are of significant biotechnological interest because they are achieved at ambient temperatures and pressures using commonly available metals at the active catalytic sites.[1,2] To understand these processes, protein-film voltammetry can be used,[3,4] where redox active proteins or enzymes are immobilised on the surface of an electrode to overcome problems caused by slow rates of macromolecular diffusion.[5] In voltammetry, a time-varying potential is used to drive electron-transfer reactions, so that the recorded experimental Faradaic current-time data directly reports on the reactions that take place.

In *voltammetric modelling*, the aim is to reconstruct the underpinning reaction process that generated the experimentally measured data.[6] For protein film voltammetry, the re-

sulting system of mathematical equations takes the form of a system of non-linear ordinary differential equations. Embedded within this system are the key parameters of interest that govern the biological redox reaction — that is, the electron transfer mechanism, reaction rates, and electrochemical potentials.[2,7] If suitable values of these parameters are chosen, it is possible to (typically computationally) solve the underlying system of equations and obtain a current-time or current potential trace describing the modelled behaviour of that system for that particular set of chosen parameters - this is known as the forwards problem. Of much greater interest, is the use of the underpinning mathematical model to recover estimates of these key parameters from experimental data - this is known as the inverse or parameter inference problem — and involves finding an optimal set of parameters that minimises a distance (or "objective") function between the experimental and simulated data. The inverse problem can also be extended using for example, Bayesian inference techniques as in this work, to obtain not only point estimates of the key parameters governing the experimental system, but also to give some measure of confidence in those estimates.

In protein film voltammetry, the solution of the inverse problem is made particularly challenging since the experimentally-measured current-time or current-potential data result not only from the Faradaic signal of interest, but also a typically very large "background" current which arises due to charging at the electrode-solution interface, compounded by the effects of uncompensated resistance. Much effort has gone into resolving this difficulty[8–10,10–12] — in the approach presented in this paper, to fully compensate for these background effects, terms modelling capacitance and resistance must be incorporated into the mathematical model, and the capacitance and resistance parameters must also be estimated as part of the inverse/inference process. This typically results in a very high-dimensional problem (in the relatively simple case considered in this paper the inference problem is in 14 dimensions) and finding an optimal solution is made very challenging by the potential presence of multiple local minima of the objective function, and these minima are often located in disparate parts of the overall parameter space. In these circumstances, it is often particularly helpful, if not essential, to make use of so-called "prior" information (typically from previous experiments or arising from known physical properties of the experimental system) to try to narrow down the parameter search space, typically by placing constraints (which are as tight as possible) on the parameter values. "As tight as possible" relates to the fact that, in solving the inference problem, only those parameter ranges which are admissible after making use of the prior information are explored. It should be noted that the Bayesian inference framework used in this paper is ideal for this problem. A key question is therefore: how can this "prior" information be obtained? In this case there are two sources. The first is any known physical constraints on the parameter from the experimental literature, or through theoretical considerations. The second is to use data arising from multiple different experimental approaches. The motivation here is that different experimental techniques are affected differently by background capacitance and resistance effects (even if those differences are sometimes subtle), and these differences can be used: (i) to devise constraints on particular parameters to narrow down the search space and render the inverse parameter inference problem tractable; and/or (ii) to exclude whole parameter sets recovered by inference algorithms that can be shown to corresponding to a local rather than the global minimum.

In this paper a multi-experiment approach is presented. To showcase its utility, the experimental system is a simple protein from Cellvibrio japonicus, referred to as *Cj*X183. This is a small domain of Cbp2D, a probable activating partner for the lytic polysaccharide monooxygenase (LPMO) produced by Cellvibrio japonicus. This LPMO catalyses the oxidative cleavage of glycosidic bonds in cellulose and such enzymes facilitate industrial biofuel production from biomass.[13] It has recently been

shown that *Cj*X183, a type-c cytochrome domain, can transfer electrons to LPMOs via the reversible redox reactivity of the heme centre,[14] with these authors making use of direct current voltammetry (DCV) measurements of the Fe3+/2+ redox chemistry of *Cj*X183 can be obtained via adsorption of purified *Cj*X183 onto a pyrolytic graphite edge electrode.

In this work widely used standard DC measurements of the electrochemistry are provided , along with two further experimental techniques — purely sinusoidal voltammetry[15] (PSV), and ramped Fourier-transform AC voltammetry[16] (r-FTACV) — to explore the same chemistry under identical experimental conditions. Lessons learned from analysis of this data is then used to develop a workflow, firstly to constrain the size of the parameter inference problem by placing constraints on individual parameters, and, secondly, to exclude whole inferred optimal parameter sets derived from one experiment where that parameter set can be shown to correspond to a local optimum by using data from a different experiment. Ultimately this approach has resulted in a rational and reproducible approach to the estimation of optimal parameters, and likely distributions of those parameters, for *Cj*X183.

This combination of the use of multiple sensitive voltametric measurement techniques and the complementary application of modern computational analysis methods allows for the delivery of a protein-film voltammetry electrochemical toolkit that enables more robust and efficient and accurate extraction of of the key biochemical parameters governing this redox system. In particular, it is possible to recover consistent parameter values, and confidence measures on those values, of reaction rates from distinct experimental runs, verifying the reproducibility of the parameter recovery process for all of the key parameters in a protein-film volatmmetry experiment.

Methods

Experimental procedures

The expression and purification of *Cj*X183 was based on protocols reported by the Hemsworth group in Leeds,[14] with several alterations. *Escherichia coli* BL21(DE3) competent cells were co-transformed with the two plasmids, pCW-*Cj*X183 and pEC86, simultaneously. Prior to inducing with IPTG the cultures were cooled to 16 °C (not 20 °C) and IPTG was added to a final concentration of 1 mM (not 0.4 mM). Post cell harvest, the cells were resuspended in 50 mL of cold lysis buffer (20 mM Tris pH 8.0, 150 mM NaCl, 10% v/v glycerol). A protease inhibitor tablet (Thermo Scientific) and 20 μL of Benzonase (2.5 U μL^{-1}) were added to the resuspended solution. The solution was placed on ice and sonicated in 30 second bursts with the same amount of cooling time in between until the cells were lysed. The lysed cell solution was centrifuged (30750 g, 4 °C, 30 min) to isolate the supernatant and imidazole added such that that imidazole concentration reached 30 mM. The protein was purified per the established protocol except with buffer A as pH 8.0, 20 mM Tris, 200 mM NaCl, and 30 mM imidazole, and buffer B which is the same except with 300 mM imidazole.

A custom-built electrochemical cell was set-up exactly as previously described for studies of wild type *Cj*X183,[14] inside an anaerobic chamber (< 40 ppm oxygen). A 10 μL aliquot of protein was pipetted onto the working electrode tip and left to adsorb until a film has formed. The cell was maintained at 5 °C (realistically between 4 and 6 °C) and all measurements were performed in pH 7, 50 mM sodium phosphate, 150 mM NaCl buffer.

Three different voltammetric methods were carried out, DCV, r-FTACV and PSV. The DCV experiments were carried out initially after protein absorption and then after r-FTACV and PSV experiments. The experiments were conducted using an Ivium potentiostat and IviumSoft control software. The potential was

cycled between -0.39 V to 0.30 V (vs the reference electrode, which is +0.239 V vs. SHE) for 4 scans with a 5 second equilibration at the start potential, using a scan rate of 30 mV s^{-1} and a potential step of 5 mV. The same method was carried out for the trumpet plot analysis with a scan rate varying from 10 to 30000 mV s^{-1}. All potentials are reported "vs. ref", using the reference value stated above.

Three successful repeats of r-FTACV were recorded for each protein film. The r-FTACV experiments were conducted using a custom potentiostat developed by collaborators at Monash University and the instrument was controlled using custom software.[16] Each r-FTACV experiment commenced with a 5 second pre-treatment at -345 mV (vs the reference electrode) and an r-FTACV potential input between -345 mV and 255 mV was applied with a scan rate of 22.35 mV s^{-1}, as well as a sinusoidal oscillation with a frequency of approximately 9 Hz and amplitude of 150 mV.[16]

Ten successful repeats of PSV were recorded for each protein film. The PSV experiments were conducted using the same potentiostat as for r-FTACV and the instrument was controlled using custom software. Each PSV experiment commenced with a 5 second pre-treatment at -45 mV (vs the reference electrode). The voltage was cycled between -344 mV to 254 mV with a frequency of approximately 9 Hz and a phase of 270°for 26.8 seconds (equivalent to 268 oscillations).

Experimental techniques: DCV, PSV and r-FTACV

In this section the relative strengths and weaknesses of the three chosen experimental techniques are summarised: Direct Current Voltammetry (DCV), Purely Sinusoidal Voltammetry (PSV), and ramped Fourier-Transform AC Voltammtry (r-FTACV).

In DCV experiments a simple cycle of a linear potential-time ramp is used as illustrated in Figure 1A. Previous studies on surface-confined proteins have shown that this simple choice of potential-time variation limits the amount of information recoverable from the current output when studying non-catalytic electron transfer processes.[5,17] This is because the total current measured in a voltammetric experiment is a sum of both Faradaic contributions and non-Faradaic "background" current which arises due to capacitive charging at the electrode-solution interface. In DCV, the low ratio of Faradaic to non-Faradaic current impedes the extraction of redox reaction parameters. However, DCV has been used extensively by electrochemists for many years, as it is relatively simple to both perform and interpret.

To overcome the shortcomings of the DCV technique, differential pulse voltammetry methods have been developed: these techniques involve applying amplitude potential pulses on a linear ramp.[16] Ramped Fourier Transform Alternating Current Voltammetry (r-FTACV) uses a large-amplitude sinusoid overlaid on top of a DCV ramp as shown in figure 1C. The large-amplitude sinusoid present in r-FTACV results in a non-linear Faradaic current response if an electroactive species is present in solution or adhered to the working electrode surface. This non-linearity can be observed in the Fourier transform of the current, where harmonic current responses ("harmonics") are observed at integer multiples of the frequency of the input sinusoid.[5,16] In contrast, an ideal double-layer capacitive background current response has a linear dependency,[18] and graphite electrodes give rise to non-Faradaic current that is only observed in the lower harmonics. Consequently, by band-selecting and inverse-Fourier transforming only higher harmonic signals, r-FTACV measurements can be analysed to extract baseline-free Faradaic current.

However, whilst r-FTACV is the most sensitive of the techniques used in this paper, its limitation is that it is very computationally expensive to simulate because of the presence of two time-scales (linear scan rate and sinewave frequency).[15] Consequently the simulation time increases as a function of the in-

put frequency, which prevents simulation of the experiments required to accurately determine rapid electron transfer rate constants. Purely sinusoidal voltammetry (PSV) can also be used to infer redox reaction parameters, as was demonstrated for an electrode-immobilised ferrocene derivative, a one-electron transfer molecule which was covalently attached to an electrode to provide approximately 80 pmol cm^{-2} surface coverage.[15] PSV can be thought of as a DCV experiment with a variable scan rate, without discontinuities at the switching potentials. The advantage of PSV is that it has just a single time constant (the potential sinusoid frequency), meaning that the computational cost is greatly reduced. This results in an experiment that can be simulated 10 times faster than r-FTACV while still containing harmonic current responses, and therefore access to purely Faradaic signals. Additionally, for a fixed number of oscillations, PSV simulation time is independent of the frequency of the potential sinusoid. Thus, while r-FTACV fitting for the ferrocene system required several hours, PSV fitting took only 45 minutes. Since solving the inverse problem (finding input parameter values that generate a simulation that is "best fit" to the experimental data) takes a significant number of simulations (tens to hundreds of thousands), this increase in simulation speed has had a significant impact on the efficiency of parameter inference.

Before describing how each of these experimental techniques can be used in complimentary fashion first to constrain the parameter search space, and then to solve the inverse parameter inference problem, the underpinning mathematical model of the experimental system is briefly delineated, along with the mathematical descriptions of the time-dependent applied potential that defines each of the three experiments.

Mathematical Model

The overall aim of this combined modelling and experimental work is to estimate the unknown reaction parameter values that give rise to the Faradaic current resulting from a surface confined redox reaction. For *Cj*X183, the mathematical model is common to all three voltammetric techniques, with the only difference being in the time-varying potential that is imposed as a boundary condition on the electrode surface. In all three techniques, the observed experimental measurement is the time-varying total current, I, which is modelled as the sum of the Faradaic current I_f, which is dependent on the rate of change of the proportion of the oxidised molecule on the electrode ψ, and the capacitive current I_c. It is assumed that the kinetics of the Faradaic process are well-described by the Butler-Volmer equations, and assume that the oxidised and reduced species do not interact with each other (i.e. a Langmuir isotherm). The effective potential E_r is the applied input potential E_{app}, minus the Ohmic drop, such that $E_r = E_{app} - IR_u$. The total expression for I_f is therefore given by

$$I_f = FA\Gamma\frac{d\psi}{dt} = (1-\psi)k_0 e^{\left(\frac{F}{RT}(1-\alpha)(E_r-E^0)\right)} - \psi k_0 e^{\left(\frac{-\alpha F}{RT}(E_r-E^0)\right)}. \quad (1)$$

The definitions all of the parameters in the mathematical model above, and in the equations defining the applied potentials below, are gathered together in table 1. The applied potential then differentiates the three different voltammetry experiments. For DCV the form is

$$E_{app}(t) = \begin{cases} E_{start} + vt, & \text{if } t \leq t_r, \\ E_{reverse} - v(t-t_r), & \text{if } t_r \leq t, \end{cases} \quad (2)$$

where E_{start} is the starting potential, from which the potential is increased to the switching potential $E_{reverse}$, and then back down again. For r-FTACV, the linear ramp (written as E_{dc}) is added to an AC waveform

$$E_{app}(t) = E_{dc}(t) + \Delta E \sin(\omega t + \eta), \quad (3)$$

where η is the phase, and ΔE the amplitude of the sinusoid. For the capacitive current I_c, the current arising from an ideal capacitor is

linearly dependent on the derivative of the effective potential. The non-linear capacitance is modelled as a 3-rd order polynomial of the form.

$$I_c = C_{dl}(1 + C_{dlE1}(E_r) + C_{dlE2}(E_r)^2 + C_{dlE3}(E_r)^3)\frac{dE_r}{dt}. \quad (4)$$

For PSV, the linear ramp is omitted to obtain

$$E_{app}(t) = E_{half} + \Delta E \sin(\omega t + \eta), \quad (5)$$

where $E_{half} = (E_{reverse}+E_{start})/2$, and all other parameters are as defined previously.

Mathematical definition of the inverse problem

The simplest form of inverse problem defines an objective function which measures the distance (typically the Euclidean or least-squares distance) between the measured experimental data and the simulated data. The solution of this inverse problem then involves finding the single set of parameter values that minimises the objective function over the whole parameter space. This is an optimisation problem and can be defined mathematically. For the case of fitting to the total current I_{data}, with N_t recorded current values, and a simulated current I_{sim} which is a function of the simulation parameters θ

$$\Phi_t(\theta) = \sqrt{\sum_{j=1}^{N_t}(I_{data}(t_j) - I_{sim}(\theta, t_j))^2}. \quad (6)$$

Parameter space is searched to find a θ that returns the minimum value of $\Phi_t(\theta)$. Alternatively, the objective function can compare data in the frequency domain, filtered to contain N harmonics $h_0, h_1 ... h_N$. An objective function is defined using least squares, where N_ω is the length of the Fourier spectrum, $\mathcal{F}$, as generated by the fast Fourier transform algorithm, and ω_k is the kth fast Fourier transform component. For the case which uses the absolute value of the Fourier spectrum,

$$\Phi_f(\theta) = \sqrt{\sum_{k=0}^{N_\omega} F^{Th}(\omega_k)(|\mathcal{F}_{data}(\omega_k)| - |\mathcal{F}_{sim}(\theta, \omega_k)|)^2}, \quad (7)$$

$$F^{Th}(\omega_k) = \sum_{n=h_0}^{h_N} F_n^{Th}(\omega_k), \quad (8)$$

F_n^{Th} is a top-hat filter (equation 9) centred on the n^{th} harmonic, and ω_{input} is the input frequency where w is the width of the window, with a maximum value of 0.5 (as the spacing between each harmonic has a value of ω).

$$F_n^{Th}(\omega_k) = \begin{cases} 1, & \text{if } \frac{n\cdot\omega_{input}}{w} \leq \omega_k \leq \frac{n\cdot\omega_{input}}{w} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

In addition to optimisation, it is possible to use an alternative, Bayesian approach to solving the inverse problem which involves defining a likelihood function which is maximised as a function of the parameters. This approach yields a distribution of possible parameter values (see, for example, figure 6). Full details of the use of Bayesian in parameter inference for voltammetric data can be found in earlier publications.[6]

There are multiple algorithmic routes to solving the inverse problem in electrochemistry, as reviewed in detail previously.[6,20] In this work the approach used to solve the minimisation problem defined in equation 6 above is to use the CMAES algorithm, which returns a single best-fit parameter vector. When taking a Bayesian approach, Markov-chain Monte Carlo (MCMC) sampling is used to yield a posterior parameter distribution, which can be analysed to determine the degree of correlation between the inferred parameter distributions. It is in the latter approach that demonstrates a more powerful methodology for electrochemical data analysis.

Optimisation

All optimisation was performed using the CMAES algorithm, as implemented in the

Table 1: Symbol Glossary

Symbol	Description
E^0	The reversible potential — the potential at which the concentration of reduced and oxidised species is equal
E^0_μ	The mean of a normal distribution of reversible potentials[19]
E^0_σ	The standard deviation (s.d.) of a normal distribution of reversible potentials[19]
k^0	The rate at which the redox reaction occurs
R_u	The portion of solution resistance that is not compensated for by the potentiostat, hence the "uncompensated resistance"
C_{dl}	The magnitude of the background current arising from linear double-layer capacitance effects
C_{dlEX}	Terms used to model non-linear capacitance effects, where X is the order number, described in the mathematical model section of the SI
Γ	The concentration of electroactive species on the surface of the electrode
ω	The frequency of the input sinusoid (for PSV and r-FTACV)
η	The phase of the Faradaic current
$C_{dl}\eta$	The phase of the capacitive current
α	The symmetry factor, which is a measure of the relative ease with which the transition state is formed in the course of the oxidative and reductive reactions

PINTS repository,[21] which returns best-fit point values for each parameter fitted. For PSV, r-FTACV and background-subtracted DCV, a least-squares objective function was constructed in either the time- or frequency-domains as appropriate. These objective functions can be found in the SI. To fit peak-potential data as observed in a trumpet plot, two objective functions were used, for the oxidative and reductive peak potentials respectively, and optimised both simultaneously.

For each fitting attempt, initial parameter values chosen randomly within the boundaries listed in table 2. Each fitting attempt was run until the change in the objective function after 200 iterations was $< 10^{-7}$. To obtain the values recorded in table 2, this process was repeated 10 times, with 10 different random initialisations, and noted the best-scoring parameter vector.

Bayesian Inference

An adaptive Markov-chain Monte-Carlo (MCMC) algorithm was used in order to obtain estimates of the posterior distribution $P(\theta|y_D)$, where y_D is some form of voltammetry data, and θ are the model parameters. More information on the algorithmic details and mathematical properties of this approach can be found in the literature, including previous work.[6] For all MCMC fitting approaches, an uninformative log-prior was used, and log-likelihood functions were as detailed in the SI. Three independent chains were initialised from best-fit estimates obtained by minimisation using the CMAES algorithm. The chains were run for 10000 samples, and the first 5000 samples were discarded as "burn-in".

Challenges in solving the inverse parameter inference problem

Solving the inverse problem for every parameter in table 1, including the three C_{dlEX} parameters, requires finding a minimum in 14 dimensions. As outlined in the Introduction, this is challenging in and of itself, but is compounded by a number of further complications. Information about the key parameters of interest (E^0, E^0_μ, E^0_σ, Γ and k_0) is contained within the Faradaic component of the total current but, for protein-film voltammetry experiments, this is swamped by background signals aris-

ing from capacitive effects and uncompensated resistance. The Faradaic signal is further obscured within the measured current by experimental measurement error, arising from processes such as shot noise, or thermal electron fluctuations.[22] The magnitude of this noise, as indicated elsewhere,[15] is proportional to the total current. Furthermore, as has been demonstrated previously,[15,19] several of the parameters in table 1 have similar effects over the potential window of interest. This manifests itself as correlations between sets of parameters, as illustrated, for example, for R_u and k_0 in figures S1 and S2. This is known as parameter compensation.[19] The effect is further compounded by the impact of the background non-Faradaic current on the parameter recovery process, which is described in more detail in the Results section. Overall, this means that the inverse problem is made very difficult by the presence of multiple local minima, often in disparate parts of the overall parameter space.

As described earlier, in this circumstance, it is particularly helpful firstly to make use of all prior information (data) to narrow down the parameter search space by placing constraints to ensure that only those parameter ranges which are admissible after making use of the prior information are included. The approach taken to derive these constraints is outlined in Figure 4 in the Results section, and full details are given in the SI. The Results section then goes on to describe in detail the workflow that allows for the use of DCV, PSV and r-FTACV data to make the overall inverse problem of parameter inference tractable, and to obtain reproducible parameter values across different experimental data sets.

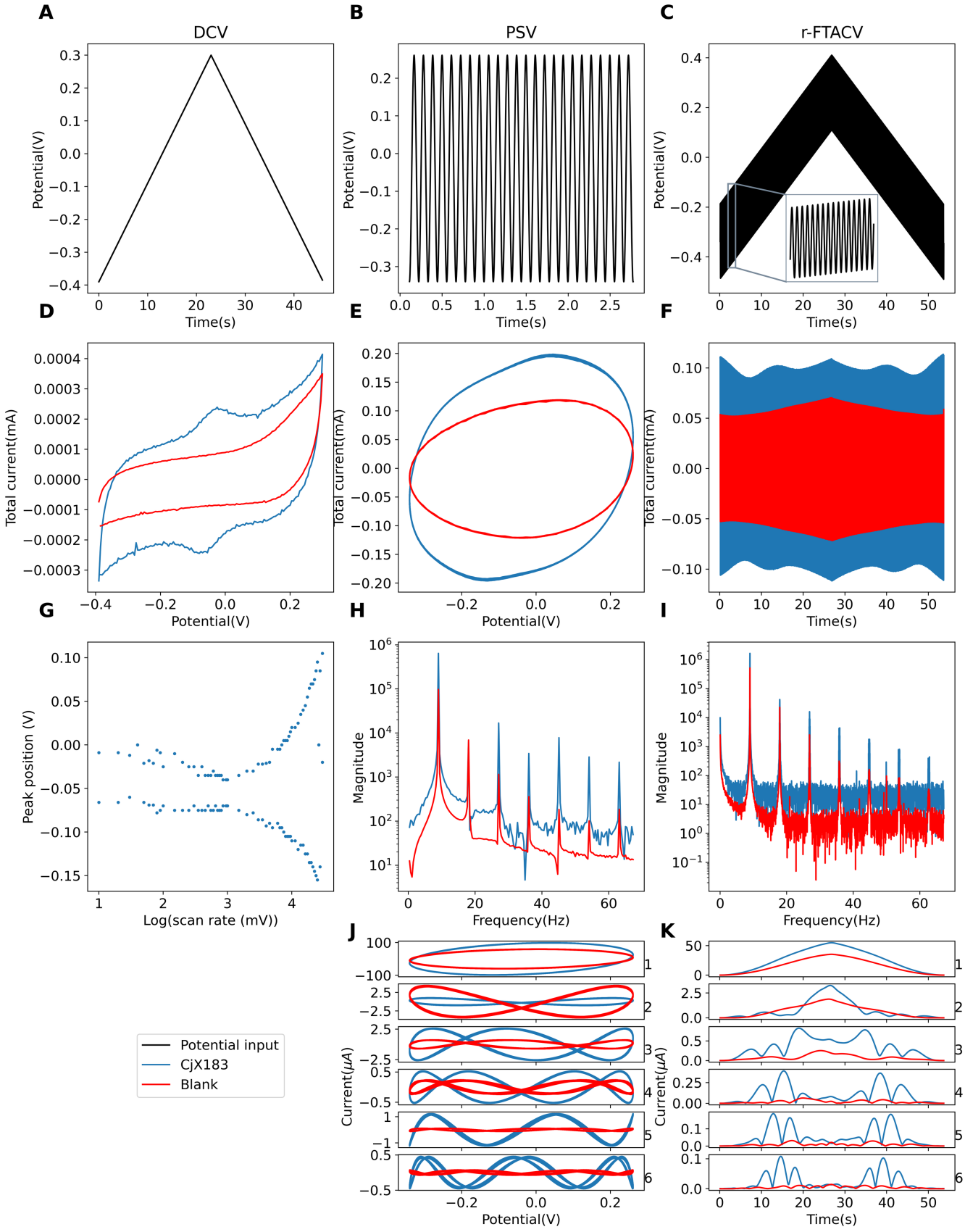

Figure 1: DCV (left), PSV (centre) and r-FTACV (right) experimental data. A-C: Potential inputs for the three experiments. D-F: Total current response of *Cj*X183 (blue) and a bare glassy-carbon electrode (orange) to the three experimental inputs. G: DCV peak position data from *Cj*X183. H: magnitude of the Fourier spectrum corresponding to harmonics 1-10 for *Cj*X183 (blue) and a bare electrode (red) for PSV and r-FTACV. I and J: PSV and r-FTACV harmonics 1-6 from *Cj*X183 (blue) and a bare electrode (red)

Results and Discussion

Preliminary Investigations

Examples of the experimental data generated by DCV, PSV and r-FTACV voltammetric experiments interrogating *Cj*X183 adsorbed to the surface of a graphite electrode are shown in figure 1, with figures 1A-C showing the potential input and the subsequent figures showing various forms of the current for the three experiments. As can be seen in figures 1D-F, comparison of experiments conducted in the presence and absence of protein ("blank" and "*Cj*X183") shows that the contribution of the Faradaic current to the total current response is significantly stronger in the DCV and r-FTACV total currents than in the PSV total current, where, to the eye, the signal attributed to the protein (the two "bumps" at $\sim \pm$ 0.18V) is almost totally overwhelmed by background current. As discussed in the introduction, one method of voltammetric analysis in the presence of large background currents is background subtraction, where the current from an unmodified electrode (i.e. purely capacitive current) is subtracted from the current of a modified electrode, to obtain a purely Faradaic current response, hypothetically free of non-Faradaic background contributions. However, as shown in figure 1, processing the *Cj*X183 experimental data in this way would not yield purely Faradaic current: the presence of protein on the electrode clearly alters the observed background current, meaning the background subtraction would only serve to bias the results.

In previous work, where parameters were inferred from a voltammetric PSV experiment interrogating ferrocene attached to a glassy carbon electrode,[15] the Faradaic-to-background current ratio was significantly higher, and consequently accurate parameters could be obtained consistently by fitting to the total PSV current. However, when fitting to the total current generated by *Cj*X183, it was determined that it was possible to obtain multiple "plausible" fits with highly dissimilar parameter values (as recorded in table S1 in the SI), with none of the proposed simulations providing a good fit to either the higher PSV or r-FTACV harmonics. Coupled with the fact that two markedly different sets of Faradaic parameters generate similar total current fits indicate that the inferred Faradaic parameters have little effect on the goodness-of-fit of the two simulations. As can be seen in figure 2, the best-fit total-current simulations are primarily a good fit to the first harmonic, which, as can be seen in figure 1, is two orders of magnitude larger than any of the higher harmonics. This is a demonstration that, within the total current of the protein-film PSV experiment, the small amount of Faradaic current arising from the biological electron transfer process is swamped by the background current contributions, as was not the case for surface-linked ferrocene. Additionally, background currents arising from a PG electrode are generally more complex than for a glassy carbon electrode, because of the roughness of the surface.

Given that fitting the total PSV current in the time domain did not provide a satisfactory result, filtering approaches were used. By inspection of the harmonics for the "blank" (protein-free) experiment in figure 1, it was judged that PSV harmonics 0-3 contained most of the capacitive current contribution, and filtered this out by band selecting harmonics 4 and above from the Fourier spectrum of the *Cj*X183 PSV current. It should be noted that this filtered data contains little information about capacitance — a detail that will be returned to in a later section.

It was also determined that the effects of thermodynamic dispersion in the model should be included, from inspection of the r-FTACV experimental harmonics, as described in previous work.[19] This can be achieved by comparing the r-FTACV harmonics against a non-dispersed simulation. Such a simulation can be obtained by fitting to PSV data, and then using the best-fit parameters to obtain an r-FTACV simulation — if the higher r-FTACV experimental harmonics are of a lower magnitude and are wider than the non-dispersed simulation, then

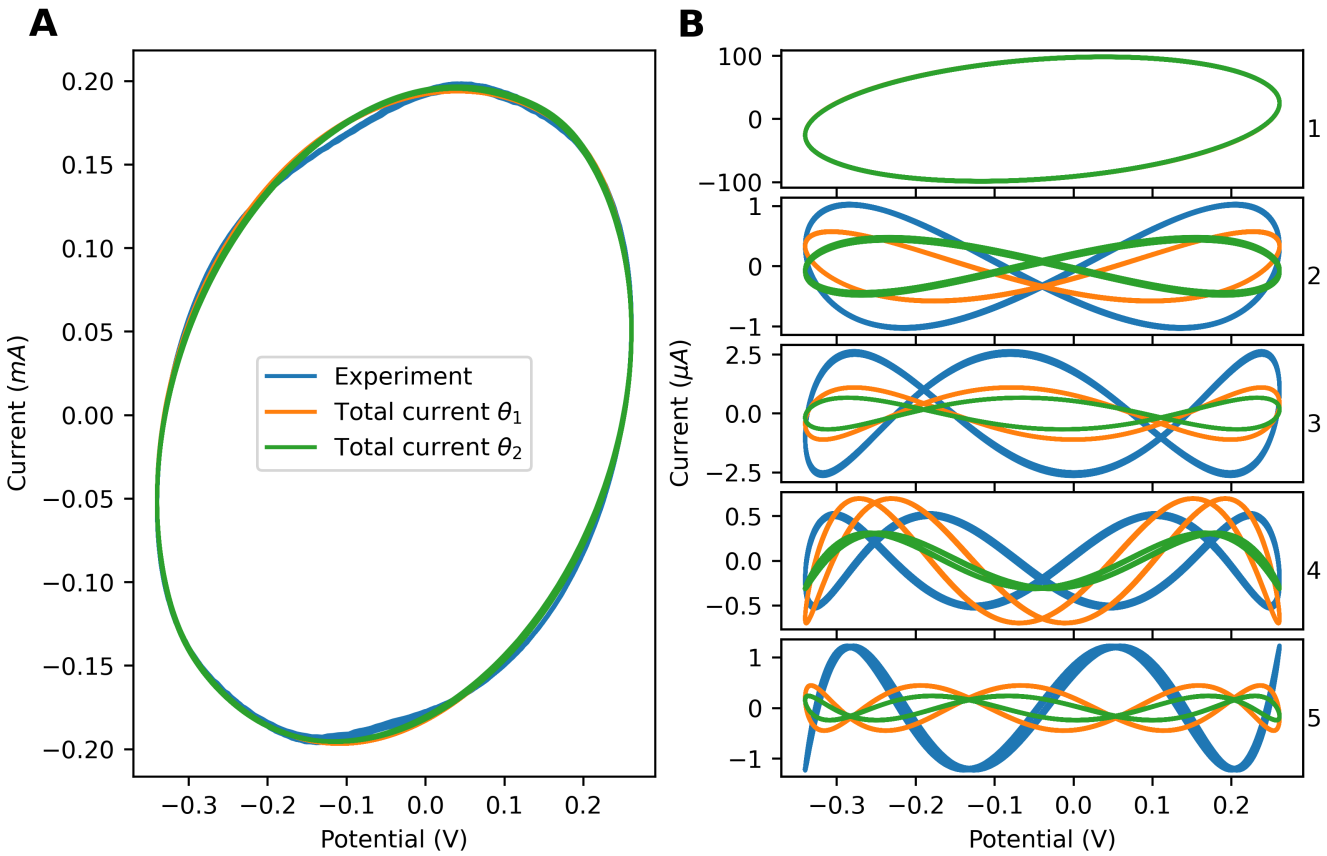

Figure 2: A: PSV total current, and two simulations with highly differing parameter values obtained by fitting to PSV total current (referred to as θ_1 and θ_2, the parameter values of which can be found in the SI in table S1). When observing the fit in harmonics 1-6 (B), it should be observed that the relatively good fit between simulation and experiment in the time-domain is not replicated in the higher harmonics.

this is an indication that thermodynamic dispersion is present. The presence of thermodynamic dispersion also implies the presence of kinetic dispersion, but in the close-to-reversible kinetic regime, kinetic dispersion does not have a significant effect on the Faradaic current, and so it was not modelled for reasons of computational efficiency[19]

After an initial parameter inference attempt, fitting to harmonics 4-10 of the *Cj*X183 current, it was determined that it was possible to obtain parameter vectors that provided a reasonably good fit to these PSV harmonics, but when compared to r-FTACV harmonics data as shown in figure 3, the fit is poor — with a mismatch between the predicted and experimental harmonic magnitudes, and implausible waveshapes — the parameters used to generate the simulation can be found in table S2 in the SI under "low k_0". The poor fit is particularly apparent in r-FTACV harmonic 4, where one single merged set of peaks is visible in the simulation while two separately resolved sets of peaks are observed in the experimental data. Based on previous studies, this type of deviation between simulation and experimental r-FTACV data is attributable to the simulation being generated by a parameter set in which the electron-transfer rate k_0 is too small. The conclusion was that that exploring a very broad region of parameter space during the modelling of PF-PSV high harmonic data is not an effective way to obtain realistic electrochemical reaction parameters. This is a direct illustration of the phenomenon of parameter compensation described earlier and illustrated in Figures S1 and S2. In these circumstances,

it is often the case that the optimisation algorithm returns a parameter set corresponding to a local (rather than the global) minimum in parameter space; in the example shown in figure 3, this is driven by the kinetic parameter.

The next section therefore summarises, described in detail for each parameter in the SI, how this problem can be avoided by using prior information and information from multiple experimental techniques to constrain the parameter search space and return a consistent and reproducible global optimum.

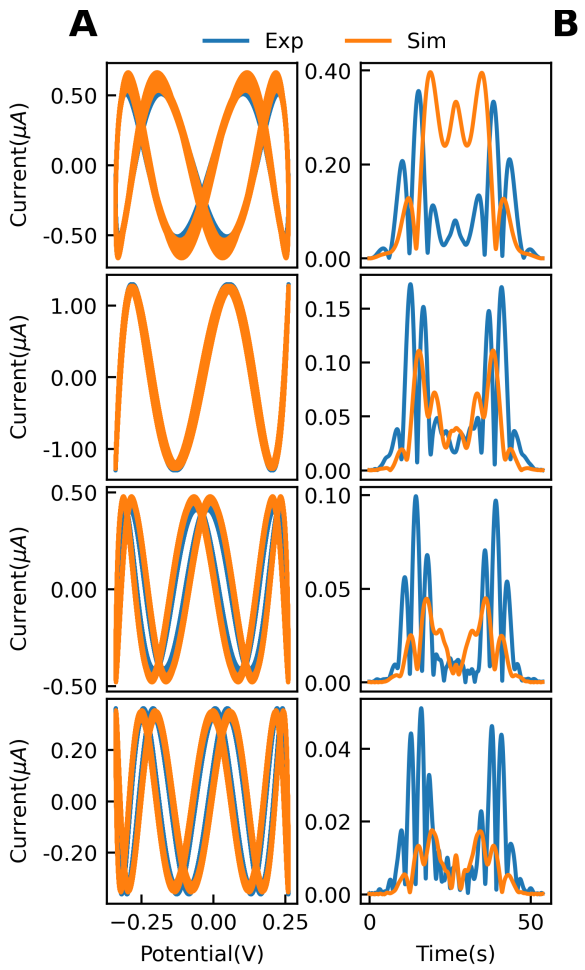

Figure 3: An example of how a relatively good agreement between PSV experimental (blue) and simulated (orange) harmonics 4-7 (A) does not translate well to an r-FTACV harmonic comparison (B), driven by a low inferred k_0 value.

Constraining the search space

From the preliminary investigation above, it became apparent that an important part of obtaining a robust fit was to constrain the parameter space in which the minimisation algorithm searches. Parameter space was constrained to prevent physically impossible parameter values, aiming to exclude the areas of space that resulted in good fits only to the filtered portion of the current from one technique, but failed to generate a good-fit to data from other techniques. This is a process that must be undertaken with great caution; it carries the obvious danger of excluding the region of parameter space in which the "true" parameter values reside. The bounds used for each parameter is shown in table 2. In the SI, the reasoning for each bound is discussed in detail in the order in which they are presented in the table, and in figure 4, this process is shown graphically.

As can be gleaned from figure 4, the process of bounding parameter space required many fitting runs to gain a deeper understanding of the inference problem. This is a situation in which the speed advantage of PSV becomes relevant. A single fitting run, returning a best-fit parameter set for PSV took approximately 45 minutes, while the same attempt with r-FTACV took several hours. Therefore, using PSV analysis, multiple model combinations can be tried in the time it takes to undertake one r-FTACV fitting attempt. As the fitting process is repeated ten times to ensure a high search coverage of parameter space, r-FTACV fitting timescales quickly become untenable.

Inferring parameters from multiple films

Using the bounds shown in table 2, it was possible infer parameters from three PSV experiments interrogating different protein preparations of *Cj*X183, where for each preparation both PSV and r-FTACV measurements were obtained. The resulting best fits are shown in figure 5, with the inferred parameters shown in table 2, where each column shows data from a different electrode film. In figures 5A-C, PSV

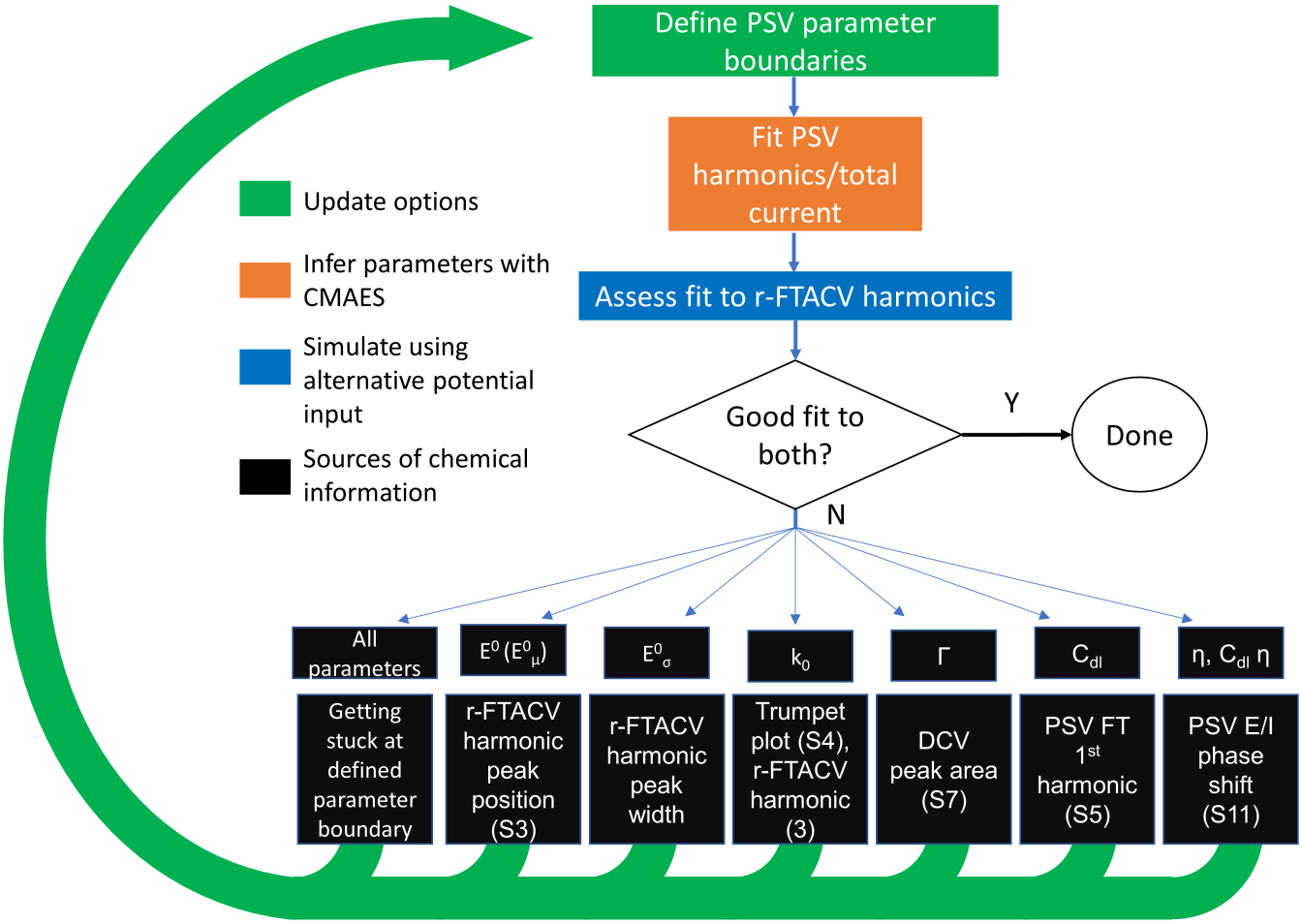

Figure 4: Flowchart representation of the fitting methodology proposed in this paper. For certain parameters the mechanism by which boundary values can be obtained has been presented graphically; these figures are written in brackets as appropriate.

harmonics 4-10 for both the experimental and simulated currents are shown, with the simulations using the parameters in table 2. In figures 5D-F r-FTACV simulations also use the input simulation parameters written in table 2, except for the input frequency and phase. For figures 5G-I, limited optimisation was performed on parameters thought likely to change as a result of so-called "film-loss" effects, which change as a function of *Cj*X183 molecules coming off the electrode over time, specifically the E^0_μ, E^0_σ and Γ parameters — the former two as the distribution of E^0 values will change as a result of molecules coming off the electrode. In addition, the sinusoidal frequency ω was fitted, which is different for r-FTACV and PSV experiments, and set the phase to zero. The altered values are written in brackets in table 2. This was done to show that only a relatively small modification in the PSV parameters is required to go from an excellent fit to the PSV harmonics to an excellent fit to the r-FTACV harmonics. Because the r-FTACV inference approach requires searching in four-dimensional parameter space, the time for a single fitting run is shorter than it would be for the full 13-dimensions searched in the PSV case. In practice, this was 2-3 hours for a r-FTACV harmonic data fitting attempt, as opposed to a predicted 10+ hours for all 13 dimensions.

The multiple comparison approach for validating the "best-fit" reaction model parameters uses both inter-technique and inter-experiment comparisons, such as the type performed in figure 5, allows the removal of spurious parameter combinations, or those that provide a good fit to only one subset of the data. Consequently,

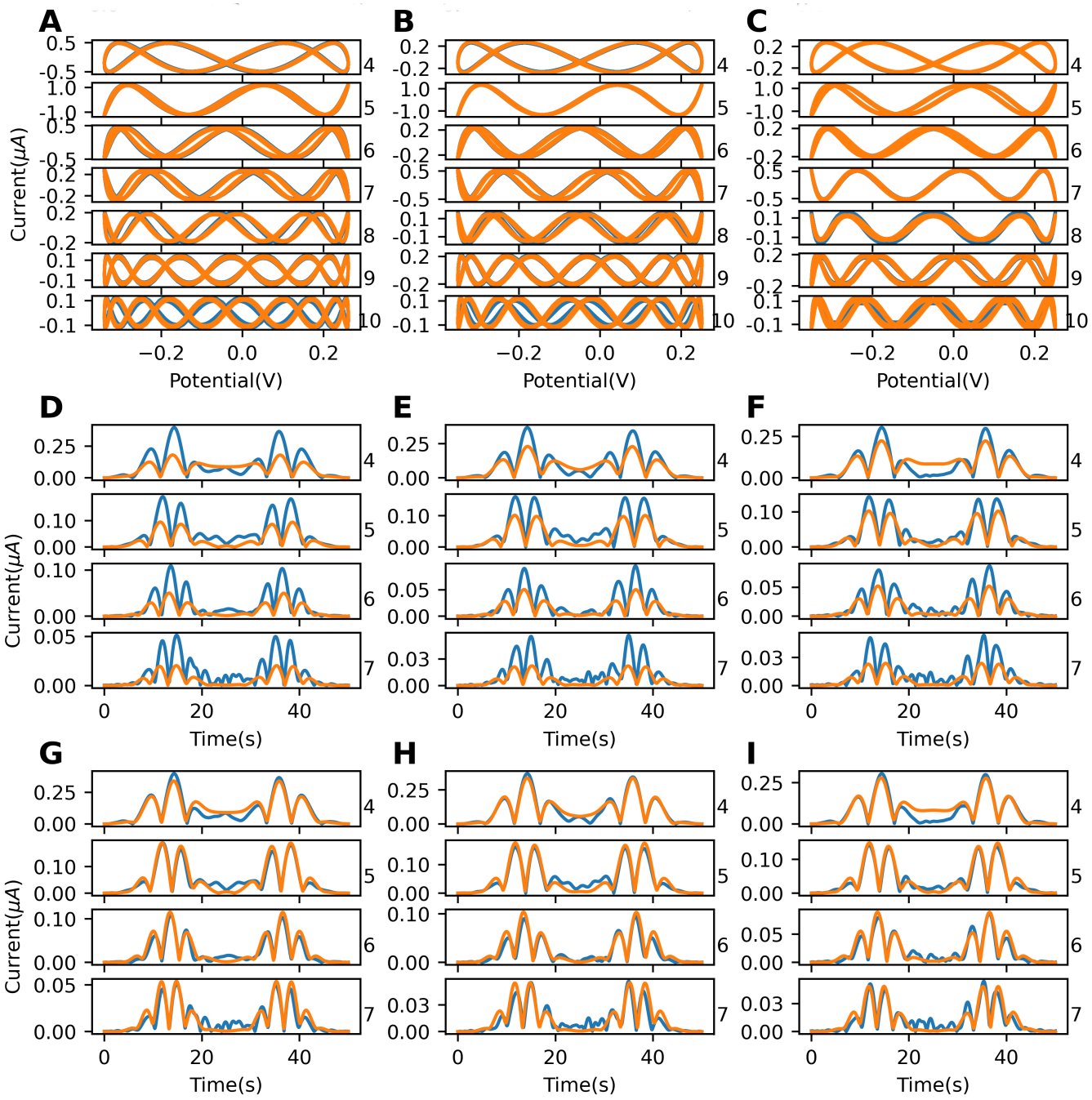

Figure 5: Best-fit simulations and data for harmonics 4-10 of PSV experiments 1-3 (figures A-C), using simulation parameters reported in table 2. Figures D-F show r-FTACV harmonics 4-7, from r-FTACV experiments 1-3 (performed using the same *Cj*X183 modified electrode as the appropriate PSV experiment), and simulations obtained using the same parameters as those used to generate the top row, except for the values of the phases (both of which were set to 0) and the input frequency (which was set to 8.96, 8.75 and 8.83 Hz respectively). Figures G-I show the same r-FTACV harmonic data as figures D-F, along with best-fit simulations, obtained using the parameters in brackets in table 2.

the fact that the values inferred from PSV data can a) with a small level of alteration provide a good fit to r-FTACV harmonics obtained during the same set of measurments and are b) consistent across experiments conducted on different days, demonstrate that the inferred values are robust. The PSV parameters can also be evaluated in light of values obtained by other analysis methods. In this study, these methods are the estimates of the kinetic parameter from trumpet plot data,[8–11,23] and surface coverage estimates from analysis of a single scan-rate DCV experiment charges calculated from peak area, shown in figures S8 and S7 respectively. With regards to the former, the value of the kinetic parameter inferred from the trumpet plot was within $60 s^{-1}$ of the value inferred from the PSV currents, and as shown in figure S8, simulating trumpet plot data with the PSV-inferred value of $\sim 172 s^{-1}$ does not significantly alter peak-position divergence point. With regards to the surface coverage, although the exact value inferred from the peak-integration technique is dependent on the subtraction approach, the values inferred from PSV are in agreement with the range of values inferred by this method.

In terms of the values that best report on the *Cj*X183 redox chemistry, because r-FTACV is more sensitive to parameter changes as demonstrated using Bayesian inference on synthetic noisy data in previous work,[15] it is the position of the authors that the values that generate the best-fit to the r-FTACV harmonics are those that best describe the properties of *Cj*X183.

The midpoint potential of a cytochrome's heme Fe^{2+}/Fe^{3+} redox couple can vary from -639 to +161 mV (vs. ref.)[24] However, in the region of physiological pH, the values for E^0 tend to range from around -49 to +11 mV (vs. ref.)[25,26] The average value extracted for the midpoint potential mean and standard deviation as part of the best fit parameters was ∼-65 mV vs reference, and as such the inferred distribution of *Cj*X183 midpoint potentials is close to or within the suggested physiological range.

When rationalising the electron transfer rate constant measured in this paper, it is important to consider that the *Cj*X183 protein under analysis is an artificially small fragment of a larger naturally occurring protein. However, the average best fit rate constant of $\sim 175\ s^{-1}$ is fully consistent with the range of values determined in previous electrochemical studies of different cytochrome proteins. For example, the electron transfer rate constant for cytochrome-c incorporated onto a calcium alginate film on a glassy carbon electrode was found to be 20.9 s^{-1},[27] the electron transfer rate constant for human, bovine and porcine cytochrome P450c17 on a PGE electrode were determined as 164, 157 and 153 s^{-1} respectively,[28] and while more varied due to changing chain length, the electron transfer rate constant for rat heart cytochrome-c adsorbed onto gold electrodes modified with self-assembled monolayer has been determined to fall around 700 s^{-1} when the length of the carbon chain on the monolayer is between 4 and 7 carbon atoms in length.[29]

Based on the geometric surface area of the working electrode (0.03 cm^2) and the width of the protein structure, by assuming spherical close packing of protein on the electrode, a theoretical maximum monolayer surface coverage of *Cj*X183 on the electrode can be calculated as 4.401 pmol cm^{-2}. The extracted best fit surface coverage parameter from the experimental data is 17.3 pmol cm^{-2}, this difference can be attributed to a non-spherical protein with better packing efficiency and a non-planar electrode surface (i.e. area greatly exceeding 0.03 cm^2 due to abrasive treatment of the graphite creating a rough surface). In general, protein film voltammetry experiments report pmol cm^{-2} coverage values.[4] What it is essential to note, however, is that the exact reported parameters are conditional on the modelling choices made. In all cases the reported α value is the value of the upper bound, and the solver converged to the value of the upper bound regardless of what this boundary was set to. This is not unexpected — as detailed in previous work, when the kinetic regime is approaching reversibility (reaction is approaching equilibrium on the

Table 2: Best fit parameters for harmonics 4 and above of PSV experiments 1, 2 and 3. The resulting simulated PSV current is shown in figures 5A-C. The same values were used to generate r-FTACV simulations shown in figures 5 D-F, and the values in brackets were used to generate the r-FTACV simulations in figures 5G-I

Parameter	Symbol	Bounds	PSV 1	PSV 2	PSV 3
Midpoint potential mean	E^0_μ (V)	[-0.1, -0.04]	-0.072 (-0.061)	-0.067 (-0.063)	-0.065 (-0.061)
Midpoint potential standard deviation	E^0_σ (V)	[1e-4, 0.06]	0.045 (0.033)	0.053 (0.036)	0.051 (0.035)
Rate constant	$k_0(s^{-1})$	[50, 500]	173.8	176.5	172.9
Surface coverage	Γ (mol cm^{-2})	[9e-12, 9e-11]	1.35e-11 (1.68e-11)	2.05e-11 (1.83e-11)	1.79e-11 (1.45e-11)
Linear double-layer capacitance	C_{dl} (F)	[1e-7, 1e-5]	9.8-6	1.0e-5	1.0e-5
1^{st} order C_{dl}	C_{dlE1}	[-0.1, 0.1]	0.014	0.079	0.095
2^{nd} order C_{dl}	C_{dlE2}	[-0.05, 0.05]	0.04	0.021	0.045
3^{rd} order C_{dl}	C_{dlE3}	[-0.05, 0.05]	-5.6e-4	-4.4e-4	-3.8e-4
Uncompensated resistance	R_u (Ω)	[0, 900]	148.7	316.8	81.5
Potential frequency	ω (Hz)	[8.56, 9.46]	9.015 (8.96)	9.015 (8.75)	9.015 (8.83)
C_{dl} phase	C_{dl} phase (rads)	[3.77 , 5.65]	4.73	4.70	4.71
Phase	Phase (rads)	[3.77 , 5.65]	4.57	4.60	4.63
Symmetry factor	α	[0.4, 0.6]	0.6	0.6	0.6

timescale of the experiment), the effect of the symmetry factor is low. The upper bound was set at 0.6 for reasons of chemical plausibility, but it is common practice to set the value of α to 0.5 under these conditions (i.e. not to fit α at all). When this was attempted, the inferred kinetic values were totally reversible (i.e. hit the upper kinetic boundary regardless of its value), which is shown in the SI. Additionally, in the SI, in table S5 and figure S12, the effect of holding the α parameter constant in the range 0.5-0.6 while fitting the other parameters is shown. When $\alpha < 0.55$ the kinetic values are total reversible — the upper k_0 bound was set at 3000 s^{-1}, and when α was below the critical value of 0.55, the solver always returned a k_0 value of 3000 s^{-1}, which is incompatible with the value obtained from the trumpet-plot analysis above.

In terms of the capacitance, the value for all three experiments is at or close to the upper bound of 1e-5F. This may indicate that the capacitance values are not physically realistic. As obtaining accurate estimates of the *Faradaic* parameters is the aim of this procedure, the primary concern is that the unrealistic capacitance estimates are not affecting the accuracy of the other inferred parameters. For example, inaccurate estimates of the background current could lead to inaccurate estimates of the level of uncompensated resistance through the Ohmic drop effect. The uncompensated resistance is in turn known to be correlated with changes in the kinetic parameter (which can also be demonstrated using Bayesian inference analysis, in a section below), and consequently poor estimates of the background current could lead to poor estimates of the kinetic value. This is why a DCV trumpet plot measurement is highly useful for setting the bounds of k_0. Indeed, the key strength of the multi-experiment approach described here is to be able to address such concerns about spurious parameter combinations. A longer and more detailed description of the capacitance modelling choices can be found in the SI.

Bayesian inference

Importantly, because of the speed at which PSV can be simulated, a much more quantitative visualisation of parameter uncertainty can be obtained. Figure 6, presents inferred distributions for key parameters for PSV experiments 1-3, along with appropriate parameters as inferred from the trumpet plot (E^0 from the trumpet plot has been graphed alongside the E^0_μ parameter as they both have the same effect on the appearance of the total current[19]). The parameters C_{dl} and α are excluded because the MCMC algorithm does not converge if the chains get stuck at an upper or lower bound. It was not practically feasible to undertake a Bayesian inference attempt for r-FTACV, because of the high computational cost of the multiple $\sim$10+ hour fitting runs required for a many-parameter MCMC. Figure 6 demonstrates that despite the above discussed issues of parameter compensation and the possibility of spurious minima, the parameter values inferred for the three separate PSV experiments are very much in the same regime, as are parameters inferred using an entirely separate experimental technique.

In figure 7, the degree of correlation using MCMC-inferred posterior distributions for PSV experiment 1 can be assessed, with the results for experiments 2 and 3 shown in figures S13 and S14. The histogram of samples for each parameter is shown along the diagonal (with the parameter indicated on the x-axis), and a scatter plot for each pair-wise combination of parameters are shown on the off-diagonals, with the $y-$ and $x-$axes indicating which parameters are being plotted together. If two parameters are totally uncorrelated, then the histogram will be a circle. An ellipse angled upwards indicates positive correlation, i.e. an increase in the value of the x-axis parameter is associated with an increase in the $y-$axis parameter, and vice versa for negative correlation. The narrower the ellipse, the stronger the degree of correlation. From the figure it is therefore clear that many parameters are correlated with each other. In particular, of

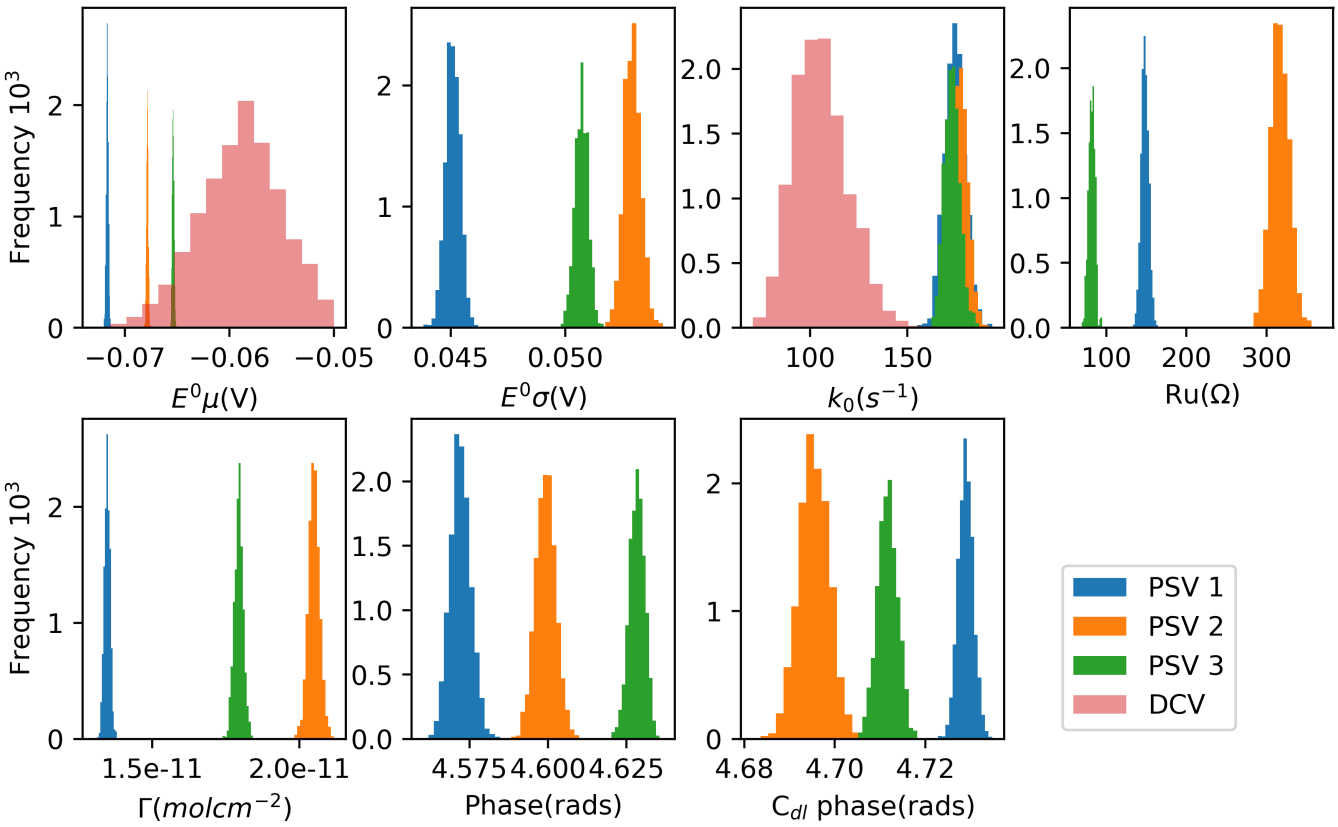

Figure 6: Inferred parameter histograms generated by pooling three independent MCMC chains, using harmonics 4 and above of the respective PSV experiments in the likelihood function, and using the parameters in table 2 as the starting point. C_{dl} and α were not included in this parameter inference approach for technical reasons mentioned in the text. The DCV histograms were as inferred from running an MCMC process on the trumpet data in figure S4, and as such only values for the parameters E^0 and k_0 were inferred.

the reaction model parameters there are correlations between k_0, the uncompensated resistance, phase, C_{dl} phase and C_{dlE2}. Thus, it is clear how challenging it is to define the electron-transfer rate. The positive correlation between E^0_σ and Γ explains that the discrepancy between the best-fit PSV and r-FTACV values is driven by these parameter compensation effects. As the two-step fits presented in figure S9 in the SI show, an increased C_{dl} value is associated with lowered E^0_σ and Γ values — indicating that the choice of C_{dl} bound (such that the returned value is relatively small) has led to slightly over-inflated predicted values for the positively correlated E^0_σ and Γ parameters. This again shows the utility of the multiple-experiment approach. For r-FTACV, this compensation between C_{dl}, E^0_σ and Γ is not of the same magnitude, which allows for the detection of the slight over-estimation described above in figure 5.

What figure 7 shows is that the choice of how to bound parameter space is not a neutral decision; because of parameter correlation, these choices affect every value returned. Consequently, although the MCMC analysis reports very high confidence in the inferred values, the exact values are contingent on the modelling choices that have been made. However, it bears repeating that this does not undermine the conclusions about the reliability of the inferred *Cj*X183 redox reaction parameters; the multi-experiment verification approach allows for verification of the inferred model parameters in a way that is not affected to the same

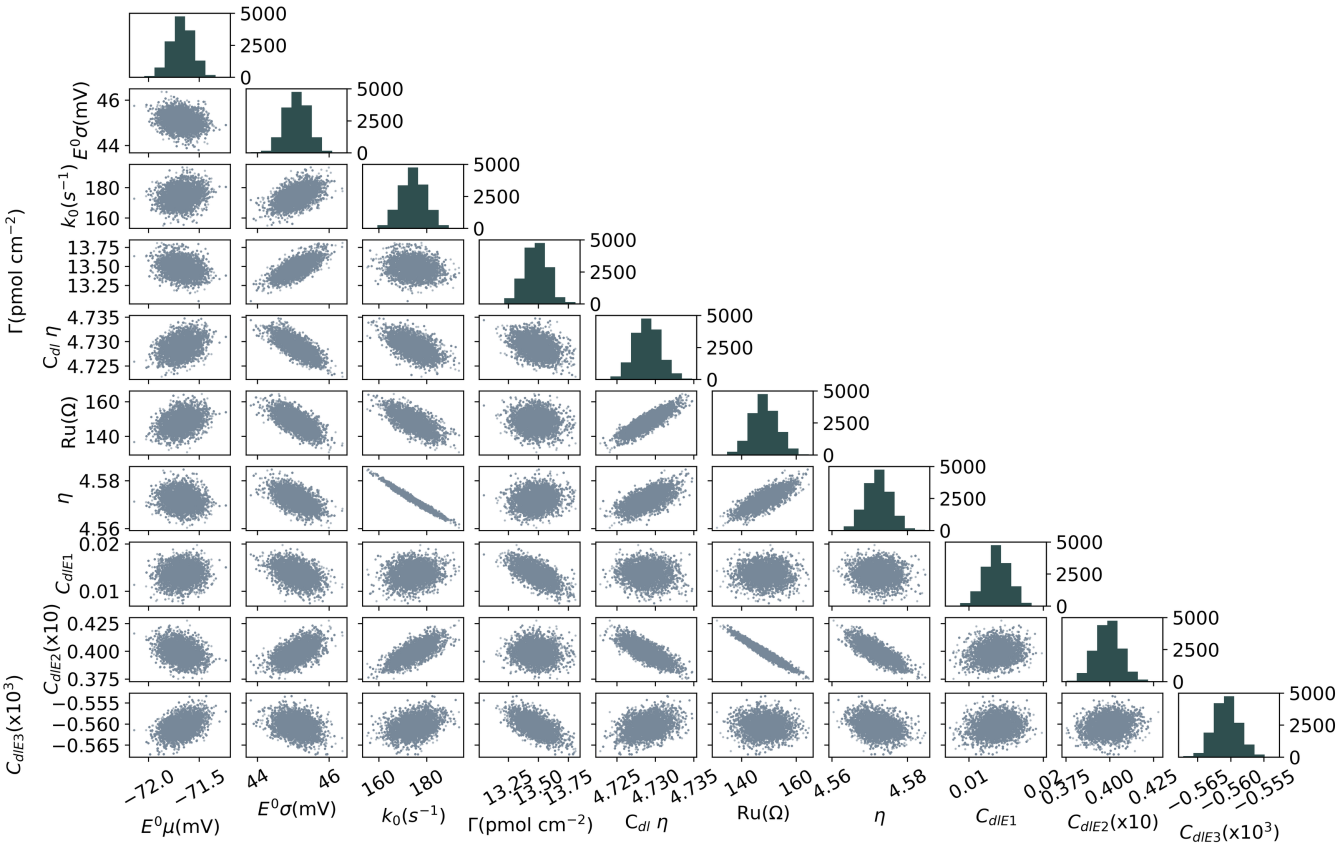

Figure 7: 2D histograms generated from the MCMC process for PSV experiment 1

extent by parameter compensations, the degree of which are peculiar to the specific data being analysed. Comparisons of experiments of the same type can be used to avoid the problem of fitting to noise in a single fitting run, and comparisons of different types of experiments provides an alternative perspective to as to spot parameter compensation effects, as with the example of C_{dl}, E^0_σ and Γ in the previous paragraph. The fact that it is possible to have this highly granular discussion about the precise values of the returned parameters is an indication of the power of this framework. The use of PSV is essential to obtain this level of understanding, as the 600,000 forward problem simulations required for a single MCMC run can be completed in just over an hour.

Conclusions

In this paper, it was demonstrated that it is possible to obtain a highly detailed picture of the electrochemistry of the single-electron redox reaction of *Cj*X183, to the extent that the parameters values inferred are consistent across multiple experiments using different preparations of the same protein. The ability to robustly determine the electrochemical properties of metalloproteins allows for tracking how these properties change as a function of different physiological conditions, providing deep insight into the bioenergetics of living systems. The multi-experiment approach, coupled with Bayesian inference, allows for increased confidence that the parameters reported are in a regime that is a good reflection of the underlying chemistry.

In previous work,[15] it was found that when inferring parameters from PSV total current-time data for a simple chemical system (a ferrocene

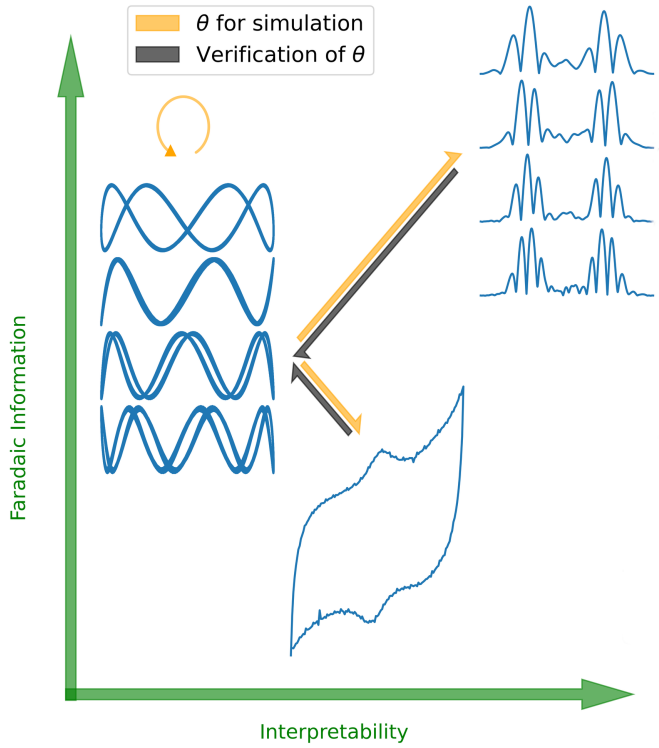

Figure 8: The three experiments analysed in this paper, direct current voltammetry (DCV) as a current vs. potential plot, purely sinusoidal voltammetry (PSV) current harmonics 4-7 vs. potential, and ramped-Fourier Transform Alternating Current Voltammetry (r-FTACV) current harmonics 4-7 vs. time. The experiments are ranked according to how interpretable they are, and the amount of Faradaic information they provide. Orange arrows indicate simulation of current using a vector of chemical parameters θ, and black arrows indicate that the results of these simulations are used to assess the goodness-of-fit.

derivative), there was one optimal parameter set which, when used to simulate a ramped experiment, also provided a good fit to r-FTACV data collected for the same electrode film. However, for the cytochrome data analysed in this paper, the signal-to-background ratio is significantly worse, and filtering approaches are required to lessen the impact of background current, which was not necessary for the analysis of ferrocene. The challenge is that this filtering discards a large quantity of Faradaic information. As a result, multiple parameter vectors that generate an equally "good fit" were found. This is exacerbated by the other weakness of PSV; because the filtered PSV current is largely uninterpretable by eye, it is not possible to a priori select one parameter vector over another. To address this problem, a protocol was developed that uses preliminary DCV analysis and visual inspection of r-FTACV to appropriately bound parameter space for subsequent PSV fitting runs. The validity of this method is indicated by the fact that this approach enables the inference of parameters from PSV high harmonic current-time data, and the "best-fit" parameter set also provides a good fit to r-FTACV data collected for the same electrode film. This process is represented graphically in figure 5. If the aim of understanding protein bio-electrochemistry by tracking changes in inferred parameters is to be realised, then parameter inference needs to be rapid, accurate and reproducible. The framework that described in this paper fulfils these three criteria. The framework laid out in this paper leverages the advantages and mitigate the disadvantages of three voltammetry techniques to provide a highly accurate protein electron-transfer analysis toolkit. In figure 8, PSV, r-FTACV and DCV methods are ranked by the "interpretability" (ease of visual inspection) and Faradaic information content of the current output of a PFV experiment. The harmonic current for the PSV and r-FTACV cases is shown as this is the form of the data analysed. In terms of Faradaic information, DCV is ranked last, because of the challenges associated with background subtraction; the highest ranked is r-FTACV, because it is highly sensi-

tive to changes in redox reaction parameters, an effect explored in previous work, and the capacitive and Faradaic current contributions can be clearly separated by filtering. In terms of interpretability, PSV is ranked last because the Faradaic contribution is difficult to discern in the oval total-current versus potential voltammogram and the appearance of the harmonics is non-intuitive; r-FTACV is ranked highest, as the effect of the distinct Faradaic parameters of interest can be determined through visual inspection. This is far from the first study to extract redox reaction parameters from PFV measurements. There is a great deal of literature, including detailed review papers,[3,4] that is related to attempts to infer redox reaction parameters from PFV data. In some sense, all of these studies require fitting a model to experimental observations. A large amount of effort and debate is expended on PFV model development; there are challenges related to distinguishing between different catalytic mechanisms,[30] along with challenges related to dispersion[31] and substrate transport.[32,33] Although the majority of published studies assume that the Faradaic reactions are well-described by Butler-Volmer kinetics (as in this work), there may be a need to incorporate Marcus theory instead.[11] Additionally, the choice of isotherm (such as the Langmuir isotherm, which is assumed in this work) is an important consideration. The key contribution is to demonstrate how this new, integrated DCV, r-FTACV and PSV measurement and analysis protocol ultimately enables Bayesian statistical analysis to be utilised to visualise the potential pitfalls in the chosen modelling approach. This is an important step forward in enabling critical evaluation of the limitations in the chosen parameter inference process. For researchers looking to use this methodology, the following "recipe" for robustly inferring parameters is proposed.

1. Collect experimental data. For each electrode functionalised with protein, it is recommended to collect PSV, r-FTACV and DCV data (the latter at different scan rates), in that order. This is because there will naturally be some film-loss as a result of consecutive experiments, and consequently the experiments have been ranked in order of how important having good signals is for the purposes of analysis.

2. Obtain estimates for E^0 and k_0 bounds by fitting trumpet plot data, a relatively simple process. This information should then be used to judge parameters inferred from analysis of PSV and r-FTACV data. You can also use MCMC at this juncture to explore the confidence in the inferred values.

3. Define boundaries for fitting PSV data. Initial boundaries should be encompass a reasonably large area of parameter space, but as a rule of thumb should not cover more than two orders of magnitude. If this scale of coverage is necessary, consider log-transformations.

4. Determine if it is feasible to fit PSV data in the time domain, without using dispersion. If inferred Faradaic parameters are highly divergent between different time-domain fitting runs, then it is recommended to fit in the Fourier domain. If this is the case, inspect the harmonics of the blank PSV data to see what portions of the Fourier domain needs to be zeroed-out, as demonstrated in figure 1.

5. Fit the form of the PSV data chosen above, using a simulation without dispersion. Using the parameters resulting from this inference process, generate a ramped simulation and compare to the r-FTACV data harmonics to assess the translatablity of the parameters. It should be reasonably clear if you are neglecting thermodynamic dispersion, as the simulated harmonics will be narrower, and will not decrease in magnitude with harmonic number to the extent observed in the experimental data. If thermodynamic dispersion is present, you should go back to fitting the PSV data accordingly. If the kinetics of the system are

irreversible/quasi-reversible then it may be worth considering kinetic dispersion as well, as discussed in previous work,[19] but this scenario has not been encountered to date.

6. Keep on comparing your PSV fits to the r-FTACV harmonics. Some other points to note:

    - If fitting to the Fourier domain, keep checking that the predicted total current simulation is not significantly greater in magnitude than the experimental current data, as shown in figure S5.
    - If a parameter is consistently hitting a defined boundary, then consider raising or lowering this bound as appropriate, unless this is outside of the realms of chemical plausibility. Beware of parameter compensation effects.
    - A good rule of thumb is that you will see a set of "good-fit parameters" multiple times in ten runs with random initialisations. Using the boundaries in table 2, the values reported were observed 2-4 times out of ten.

7. Choosing which parameters to report is something of a personal choice — the rationale was that the inferred parameters for the three experiments should be in the same regime while providing a good fit to each. For future work, the authors intend to extend the repertoire of techniques to include higher frequency PSV experiments, square wave voltammetry and electrochemical impedance spectroscopy, and to move towards systems that have more complex chemistries, including multiple electron-transfer reactions and catalytic processes.

References

(1) Beaton, S. E.; Evans, R. M.; Finney, A. J.; Lamont, C. M.; Armstrong, F. A.; Sargent, F.; Carr, S. B. The structure of hydrogenase-2 from Escherichia coli: implications for H2-driven proton pumping. *Biochemical Journal* **2018**, *475*, 1353–1370.

(2) Fourmond, V.; Wiedner, E. S.; Shaw, W. J.; Léger, C. Understanding and design of bidirectional and reversible catalysts of multielectron, multistep reactions. *Journal of the American Chemical Society* **2019**, *141*, 11269–11285.

(3) Gulaboski, R.; Mirčeski, V.; Bogeski, I.; Hoth, M. Protein film voltammetry: electrochemical enzymatic spectroscopy. A review on recent progress. *Journal of Solid State Electrochemistry* **2012**, *16*, 2315–2328.

(4) Léger, C.; Bertrand, P. Direct electrochemistry of redox enzymes as a tool for mechanistic studies. *Chemical Reviews* **2008**, *108*, 2379–2438.

(5) Adamson, H.; Bond, A. M.; Parkin, A. Probing biological redox chemistry with large amplitude Fourier transformed ac voltammetry. *Chemical Communications* **2017**, *53*, 9519–9533.

(6) Gavaghan, D. J.; Cooper, J.; Daly, A. C.; Gill, C.; Gillow, K.; Robinson, M.; Simonov, A. N.; Zhang, J.; Bond, A. M. Use of Bayesian inference for parameter recovery in DC and AC Voltammetry. *ChemElectroChem* **2018**, *5*, 917–935.

(7) Fourmond, V.; Léger, C. Modelling the voltammetry of adsorbed enzymes and molecular catalysts. *Current Opinion in Electrochemistry* **2017**, *1*, 110–120.

(8) Stan, R. C.; Kros, A.; Akkilic, N.; Sanghamitra, N. J.; Appel, J. Direct wiring of the azurin redox center to gold electrodes investigated by protein film voltam-

metry. *Journal of Electroanalytical Chemistry* **2017**, *787*, 14–18.

(9) Fleming, B. D.; Zhang, J.; Elton, D.; Bond, A. M. Detailed analysis of the electron-transfer properties of azurin adsorbed on graphite electrodes using dc and large-amplitude Fourier transformed ac voltammetry. *Analytical Chemistry* **2007**, *79*, 6515–6526.

(10) Armstrong, F. A.; Heering, H. A.; Hirst, J. Reaction of complex metalloproteins studied by protein-film voltammetry. *Chemical Society Reviews* **1997**, *26*, 169–179.

(11) Armstrong, F. A. Insights from protein film voltammetry into mechanisms of complex biological electron-transfer reactions. *Journal of the Chemical Society, Dalton Transactions* **2002**, 661–671.

(12) Munge, B.; Das, S. K.; Ilagan, R.; Pendon, Z.; Yang, J.; Frank, H. A.; Rusling, J. F. Electron transfer reactions of redox cofactors in spinach photosystem I reaction center protein in lipid films on electrodes. *Journal of the American Chemical Society* **2003**, *125*, 12457–12463.

(13) Leggio, L. L.; Simmons, T. J.; Poulsen, J.-C. N.; Frandsen, K. E.; Hemsworth, G. R.; Stringer, M. A.; Von Freiesleben, P.; Tovborg, M.; Johansen, K. S.; De Maria, L., et al. Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. *Nature Communications* **2015**, *6*, 1–9.

(14) Branch, J.; Rajagopal, B. S.; Paradisi, A.; Yates, N.; Lindley, P. J.; Smith, J.; Hollingsworth, K.; Turnbull, W. B.; Henrissat, B.; Parkin, A., et al. C-type cytochrome-initiated reduction of bacterial lytic polysaccharide monooxygenases. *Biochemical Journal* **2021**, *478*, 2927–2944.

(15) Lloyd-Laney, H. O.; Yates, N. D.; Robinson, M. J.; Hewson, A. R.; Firth, J. D.; Elton, D. M.; Zhang, J.; Bond, A. M.; Parkin, A.; Gavaghan, D. J. Using Purely Sinusoidal Voltammetry for Rapid Inference of Surface-Confined Electrochemical Reaction Parameters. *Analytical Chemistry* **2021**, *93*, 2062–2071.

(16) Bond, A. M.; Duffy, N. W.; Guo, S.-X.; Zhang, J.; Elton, D. Changing the look of voltammetry. *Analytical Chemistry* **2005**, *77*, 186–A.

(17) Armstrong, F. A.; Camba, R.; Heering, H. A.; Hirst, J.; Jeuken, L. J.; Jones, A. K.; Léger, C.; McEvoy, J. P. Fast voltammetric studies of the kinetics and energetics of coupled electron-transfer reactions in proteins. *Faraday discussions* **2000**, *116*, 191–203.

(18) Bard, A. J.; Faulkner, L. R., et al. Fundamentals and applications. *Electrochemical methods* **2001**, *2*, 580–632.

(19) Lloyd-Laney, H. O.; Robinson, M. J.; Bond, A. M.; Parkin, A.; Gavaghan, D. J. A Spotter's Guide to Dispersion in Non-Catalytic Surface-Confined Voltammetry Experiments. *Journal of Electroanalytical Chemistry* **2021**, 115204.

(20) Bieniasz, L.; Speiser, B. Use of sensitivity analysis methods in the modelling of electrochemical transients: Part 3. Statistical error/uncertainty propagation in simulation and in nonlinear least-squares parameter estimation. *Journal of Electroanalytical Chemistry* **1998**, *458*, 209–229.

(21) Clerx, M.; Robinson, M.; Lambert, B.; Lei, C. L.; Ghosh, S.; Mirams, G. R.; Gavaghan, D. J. *Journal of Open Research Software* **2019**,

(22) Gao, R.; Edwards, M. A.; Harris, J. M.; White, H. S. Shot noise sets the limit of quantification in electrochemical measurements. *Current Opinion in Electrochemistry* **2020**, *22*, 170–177.

(23) Laviron, E. General expression of the linear potential sweep voltammogram in

the case of diffusionless electrochemical systems. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry* **1979**, *101*, 19–28.

(24) Mao, J.; Hauser, K.; Gunner, M. How cytochromes with different folds control heme redox potentials. *Biochemistry* **2003**, *42*, 9829–9840.

(25) Hendler, R.; Sidhu, G. A new high potential redox transition for cytochrome aa3. *Biophysical journal* **1988**, *54*, 121–133.

(26) Jünemann, S.; Wrigglesworth, J. M.; Rich, P. R. Effects of decyl-aurachin D and reversed electron transfer in cytochrome bd. *Biochemistry* **1997**, *36*, 9323–9331.

(27) Jian, S.; Liu, X.; Sun, H.; Hou, S. The electrochemical studies of cytochrome c incorporated in 3D porous calcium alginate films on glassy carbon electrodes. *RSC Advances* **2014**, *4*, 6165–6172.

(28) Johnson, D. L.; Conley, A. J.; Martin, L. L. Direct electrochemistry of human, bovine and porcine cytochrome P450c17. *Journal of molecular endocrinology* **2006**, *36*, 349–359.

(29) Wei, J.; Liu, H.; Niki, K.; Margoliash, E.; Waldeck, D. Probing electron tunneling pathways: Electrochemical study of rat heart cytochrome c and its mutant on pyridine-terminated SAMs. *The Journal of Physical Chemistry B* **2004**, *108*, 16912–16917.

(30) Hirst, J. Elucidating the mechanisms of coupled electron transfer and catalytic reactions by protein film voltammetry. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **2006**, *1757*, 225–239.

(31) Léger, C.; Jones, A. K.; Albracht, S. P.; Armstrong, F. A. Effect of a dispersion of interfacial electron transfer rates on steady state catalytic electron transport in [NiFe]-hydrogenase and other enzymes. *The Journal of Physical Chemistry B* **2002**, *106*, 13058–13063.

(32) Merrouch, M.; Hadj-Saïd, J.; Léger, C.; Dementin, S.; Fourmond, V. Reliable estimation of the kinetic parameters of redox enzymes by taking into account mass transport towards rotating electrodes in protein film voltammetry experiments. *Electrochimica Acta* **2017**, *245*, 1059–1064.

(33) Reda, T.; Hirst, J. Interpreting the catalytic voltammetry of an adsorbed enzyme by considering substrate mass transfer, enzyme turnover, and interfacial electron transport. *The Journal of Physical Chemistry B* **2006**, *110*, 1394–1404.